SOME RESULTS IN BHATTACHARYYA DISTANCE-BASED LINEAR DISCRIMINATION AND IN DESIGN OF SIGNALS

A Thesis Submitted
in Partial Fulfilment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

by
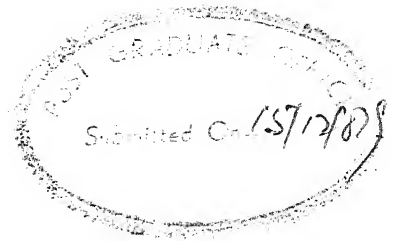
GOPAL CHAUDHURI

to the

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY KANPUR
DECEMBER, 1989

To

The loving memory
of my parents

CERTIFICATE

Certified that this work, entitled, "SOME RESULTS IN BHATTACHARYYA DISTANCE-BASED LINEAR DISCRIMINATION AND IN DESIGN OF SIGNALS" by Gopal Chaudhuri, has been carried out under our supervision and has not been submitted elsewhere for a degree.

P.R.K. Rao
Professor
Department of Electrical Engineering
Indian Institute of Technology
Kanpur

J.D. Borwanker
Professor
Department of Mathematics
Indian Institute of Technolo
Kanpur

December 15 , 1989

ACKNOWLEDGEMENTS

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Some Notations and Symbols

ABS	absolute value
E	expectation operator
exp	exponential
sup	supremum
max	maximum
min	minimum
$\lambda_{max}(A)$	maximum eigen value of A
$\lambda_{min}(A)$	minimum eigen value of A
$N_n(.,.)$	n-dimensional multivariate normal distribution
$\overline{\lim}$	upper limit
$\underline{\lim}$	lower limit
Cov	covariance
$\Psi'(x)$	derivative of $\Psi(x)$ with respect to x
p.d.	positive definite
$\Phi(x)$	standard normal distribution function
$\underset{\sim}{\alpha}_*$	optimal $\underset{\sim}{\alpha}$, also sometimes denoted by $\hat{\underset{\sim}{\alpha}}$
$\underset{\sim}{\alpha}^*$	complex conjugate-transpose of $\underset{\sim}{\alpha}$
$\doteq$	approximately
$\overset{\Delta}{=}$	defined by
$|A|$	determinant of a matrix A
$\sim$	(at bottom) a vector
$\top\!\top$	statistical population
A'	transpose of a matrix A

vi

e.v.	eigen vector
$\bar{\underset{\sim}{\alpha}}$	complex conjugate of $\underset{\sim}{\alpha}$
[a,b]	a closed interval
L_2	space of square integrable functions on a specified interval
ε	belongs to
Re(z)	real part of z
Im(z)	imaginary part of z
$|x|$	absolute values of x
Σ	summation sign
$\rightarrow$	converges to
d(x,y)	distance between x and y
δ_D	Dirac delta function
δ_{nm}	Kronecker delta
$\Longrightarrow$	implies
X $\sim$	X is distributed as
$\Longleftrightarrow$	if and only if
$\ni$	such that

SYNOPSIS

GOPAL CHAUDHURI
DEPARTMENT OF MATHEMATICS

SOME RESULTS IN BHATTACHARYYA DISTANCE-BASED LINEAR DISCRIMINATION AND IN DESIGN OF SIGNALS

1. DR. J.D. BORWANKER
2. DR. PRK RAO

DECEMBER, 1989

In many practical situations it is of interest to classify a normal time series as belonging to one or the other of two categories described by two hypotheses. The admissible procedure for classification provided by the Neyman-Pearson theory as well as the Bayes' rule are based on the likelihood ratio . In the case of unequal covariance matrices this likelihood ratio depends on a quadratic function of observations. Unfortunately , the distribution theory pertaining to the quadratic part of this classification rule is extremely complicated . It involves the weighted sum of non-central chi-square random variables so that computing error rates resulting from its use seems difficult . Hence the usual approach has been to consider a linear procedure. In this work we study optimal classification rules based on linear statistics which maximize the Bhattacharyya distance.

The present thesis is divided into seven chapters. A brief review of the previous work done in the area of two-group classification relevent to our discussion, and a chapter-wise outline of this thesis are given in Chapter I.

Chapter II provides some definitions and elementary results needed in subsequent chapters of the present work.

Chapter III attempts at making a systematic study of the optimal classification rules based on linear statistics which maximize the Bhattacharyya distance (B-D) in the case when the observed process is discrete in time. Both stationary and non-stationary cases have been considered. For an arbitrary time series, we have to solve iteratively an implicit equation in order to get our desired linear discriminant function (LDF); we have given a simple method for ascertaining an interval of convergence of the iteration process. Linear procedures based on the B-D belong to the Anderson-Bahadur admissible class for a proper choice of the cut-off point. Some special categories of problems where the mean vectors and the covariance matrices are of specific kind have also been studied here. We compare the performance of our LDF with that of some other LDFs considered in the literature. The comparison with the quadratic discriminant function due to the Bayes' criterion is considered when covariance matrices are proportional. These comparisons result in the conclusion that the distance of our interest is worth-considering. A one-to-one correspondance between the two classes of linear procedures ---- one due to our criterion of maximizing the B-D

and another of minimizing the total probability of misclassification subject to a linear relationship between the two types of errors --- has been established. Under certain regularity conditions, a compact form of the LDF is obtained for a covariance stationary time series when the sample size is large. Some illustrative examples satisfying the regularity conditions are given and it is shown that errors of misclassification tend to zero asymptoticall

Chapter IV deals with the continuous time series. It is shown that finding the optimal LDF amounts to solving an integral equation of Fredholm type. We are able to obtain a compact form of the LDF in the case when the time series is covariance stationary with the observation interval infinite.

Chapter V is devoted to designing of signals. It is shown that even in the simple case when the two types of errors are made equal we are unable to obtain an explicit expression for the optimum signal. An analytical solution for the optimal signal however is available only through a bound on the total probability of misclassification.

As complex normal processes are of interest in many applied areas ; in Chapter VI, all the major results of the preceeding chapters have been extended to the case where the underlying process is complex valued. Here too, some special categories of problems of practical interest have been discussed.

The concluding Chapter VII includes some suggestions for further investigation in this interesting field.

CHAPTER I

INTRODUCTION

1.1 PROBLEM

The problem of classification, also known as discrimination
or identification in statistical literature, arises when an
observation is to be classified as coming from one of several
categories or populations characterized by their respective
probability distributions.   In many cases it can reasonably
be assumed that there is a finite number of populations
from which the observation could have come.

The problem of classification may be viewed as a problem
of statistical decision functions ([3], Chapter 6). We have
a number of hypotheses : each hypothesis is that the distribution
of the observation is a given one. We must accept one of
these hypotheses and reject the others.   If we are concerned
with only two populations, we have an elementary problem
of testing one simple hypothesis against another simple
one.

In developing a classification rule, distributional assumptions may be :

1. that the probability distributions of the observation under the populations are completely known,

2. that the functional form of the distributions is known but the parameters are unknown, or

3. that nothing whatever is known about the distributions.

In this work we consider the classification problem in the case of two multivariate normal distributions with different mean vectors and covariance matrices. We assume all parameters known.

The two distributions of the random vector $\underset{\sim}{x}$ of n components are denoted by $N_n(\underset{\sim}{\mu}_1, R_1)$ and $N_n(\underset{\sim}{\mu}_2, R_2)$, where $\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2$ are the mean vectors and R_1 and R_2 are the covariance matrices of the first and second populations, respectively ; the density of the jth distribution (j = 1,2) is

$$p_j(\underset{\sim}{x}) \equiv \frac{1}{(2\pi)^{\frac{n}{2}} |R_j|^{\frac{1}{2}}} \exp\{-\tfrac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu}_j)' R_j^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_j)\} \qquad (1.1.1)$$

The theoretically best procedures for classification (or, alternatively, for testing the null hypothesis of one distribution against the alternative hypothesis of the other distribution) are based on the likelihood ratio $p_2(\underset{\sim}{x})/p_1(\underset{\sim}{x})$; one classifies into the first population if

this ratio (for a given observation $\underset{\sim}{x}$) is less than a
constant and into the second otherwise. If $R_1 = R_2$,
the likelihood ratio depends on a linear function of
$\underset{\sim}{x}$ (called the Fisher's discriminant function), but if
$R_1 \neq R_2$ the ratio depends on a quadratic function of $\underset{\sim}{x}$ ([21]).
In particular, in the univariate case , the logarithm of
the likelihood ratio is

$$\ln \frac{\sigma_1}{\sigma_2} + \frac{1}{2}(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2})x^2 - (\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2})x + \frac{1}{2}(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2}) \qquad (1.1.2)$$

where $R_1 = \sigma_1^2$ and $R_2 = \sigma_2^2$. If $\sigma_2^2 > \sigma_1^2$, the coefficient
of x^2 is positive, and the set of x's for which (1.1.2)
is less than a constant is a finite interval. The procedure
is to classify an observation as coming from the first
population if it falls in this interval and as from the
the second if it falls outside (i.e. if the observation is
sufficiently small or large). In the bivariate case, the
regions are defined by conic sections ; for example, the
region of classification into one population might be the
interior of an ellipse or the region between two hyperbolas.
In general,the regions are defined by means of a quadr -
atic function of the observations which is not necessarily
a positive definite quadratic form. These procedures
depend very much on the assumption of normality and in
particular on the shape of the normal distribution relative.

to its center. For instance, in the univariate case cited
above, the region of classification into the first population
is a finite interval because the density of the first
population falls off in either direction more rapidly than
the density of the second since its standard deviation is
smaller.

In a situation where the two populations are centered
around different points and have different patterns of
scatter, and where one considers multivariate normal
distributions to be reasonably good approximations for
these two populations, one may want to divide the sample
space into two regions of classification by some simple
curve or surface. The simplest one is a line or a hyperplane ;
the procedure may then be termed linear. The formal definition
of a linear procedure is given in Chapter III. It is useful
to consider the linear procedures as the distributional
problems associated with the quadratic classification function
are very complicated ([59]) (in actual practice the error rates
are found through bounds, see ([63]) and even its implementation
is difficult. We study an optimal classification rule based
on linear statistics which maximizes the Bhattacharyya distance.
We define the Bhattacharyya distance in Chapter II.

One naturally asks : why did we pick up the Bhattacharyya distance for our analysis ? The points which motivated us to examine the consequence of maximizing the Bhattacharyya distance for linear functions are the following :

1. In the context of control theory, Schweppe ([26]) made the following remark : ''readers who subscribe to the theory that 'the best answer is the simplest answer ' may decide that the Bhattacharyya distance is superior to the Kullback—Leibler distance '' .

2. In the study of the problem of signal selection (we define it shortly) when the covariances are equal, the Bhattacharyya distance has been employed successfully (see [26,48]).

3. If one uses Bayes criteria for classification and attach equal costs to each type of misclassification, then it is shown by Matusita ([39]) that the total probability of misclassification is majorized by $\exp \{-B\}$, where B denotes the Bhattacharyya distance.

4. In the case of equal covariances, the maximization of Bhattacharyya distance yields the Fisher's discriminant function.

A problem closely related to the problem of classification is the following. In our analysis the observed random vector $\underset{\sim}{X}$ constitutes a time series (a time series is a collection of observations made sequentially in time). A fairly general model ([5]) of a time series can be written as

$$X(t) = \mu(t) + n(t)$$

where $\mu(t)$ is a completely deterministic process and $n(t)$ is a stochastic process. They are sometimes called "signal" and "noise" process respectively. We assume that $\underset{\sim}{X}$ is an observation on the stochastic process $\{X(t), t \varepsilon T\}$.

Let

$$X(t) = \begin{cases} \mu(t) + n_1(t), \text{ under } H_1 \\ \\ n_2(t), \text{ under } H_2, \end{cases}$$

where $n_j(t)$ are normal processes.

Our object is to minimize the Bayes risk or, if we attach equal cost to the two types of errors, to minimize the total probability of misclassification. For a given $\underset{\sim}{\mu}$, this probability will also be a function of $\underset{\sim}{\mu}$ and one naturally asks : which is the signal vector that minimizes the total probability of error due to the Bayes optimal classification rule subject to the condition $\underset{\sim}{\mu'}\underset{\sim}{\mu} = 1$? This problem is referred to as the "design of signal" ([48]). The direct minimization of the total probability of error is extremely difficult even when $R_1 = R_2$. We thus adopt some signal selection criterion for our purpose. We study the case when $R_1 \neq R_2$.

Besides this interesting area of signal selection, there are a number of practical problems which reduce to classifying

a realization of a normal stochastic process as belonging
to one or the other of two categories, e.g. discriminating
between seismic records originating from earthquakes and
those originating from nuclear explosions.   Applications
of time series discriminant analysis are not limited to
the physical sciences. The classification of individuals
using recorded brain waves is a potentially important
application in medicine.

1.2   REVIEW OF PREVIOUS WORK

The origin of classification problem is fairly old
and its development reflects the same broad phases as
those of general statistical inference, viz., a Pearsonian
stage followed by Fisher, Neyman-Pearsonian and Waldian stages.

In early work , the classification problem was not precisely
formulated and often considered as the problem of testing the
equality of two or more distributions. Various test statistics
were proposed in order to measure the distance between two
populations. It was Pearson who first proposed one such
statistic and termed it as the "coefficient of racial
likeness'' to ascertain the statistical distance between
two samples (K. Pearson, 1926, [46]). Dissatisfaction with
Pearson's coefficient led P.C. Mahalanobis to propose the
D^2-statistic as an alternative (Mahalanobis, 1936, [34]).

This was first successfully applied to discrimination problems in anthropological studies among others. The D^2-statistic has become widely used because it is an actual measure of metric distance between population centroids rather than primarily a criterion for testing the null hypothesis of zero distance. Here by population we mean the normal populations with equal covariance matrices.

The first published accounts of what we now know as discriminant analysis in the strictest sense were in craniometrics by Barnard (1935, [8]) and Martin (1936, [35]). These authors had the procedures suggested to them by Sir Ronald Fisher, who is rightfully given the credit for developing the discriminant function technique. In 1936 ([17]) Fisher formally proposed the linear discriminant function as a solution to a practical problem of achieving optimal seperation of two species of plants using a number of dependent variables. The motivation for the use of the linear discriminant function in multivariate populations came from Fisher's own idea in the univariate case.

For the univariate case he suggested a rule which classifies an observation x into the ith univariate population if

$$|x-\bar{x}_i| = \min \{|x-\bar{x}_1|, |x-\bar{x}_2|\} \quad , \quad (i = 1,2)$$

where $\bar{x}_i$ is the sample mean based on a sample of size n_i from the ith population. For an n-component observation vector $(n > 1)$, Fisher reduced the problem to the univariate one by considering an optimum linear combination of the n-components. He obtained it by maximizing the ratio of the difference of the expected values of a linear combination under the two populations to its standard deviation. He then used his univariate discrimination method with this optimum linear combination of components as the random variable.

The next stage of development of discriminant analysis was influenced by Neyman and Pearson's pioneering fundamental work (1936,[41]) in the theory of statistical inference. The Fisher's classification aspect received mathematical validity when Welch (1939, [69]) showed that the identification aspect of Fisherian discriminant function was essentially an application of the Neyman-Pearson likelihood ratio principle. Actually he derived the form of Bayes rule for discriminating between two known multivariate populations with the same covariance matrix ; he illustrated the theory with multivariate normal populations. This example was also taken up by Wald (1944, [66]) when the parameters were unknown ; replacing the unknown parameters by their respective maximum likelihood estimates he studied the distribution of his proposed test statistic.

The problem of classification of an observation into two univariate normal populations with different variances

was studied by Cavalli (1945, [13]) and Penrose (1947, [47]). The multivariate analog was treated by Smith (1947, [61]). He suggested Monte Carlo methods for computing error rates. He gave an example of bivariate normal distributions.

The rigorous mathematical treatment of the classification problem was put forth in a series of papers by C.R. Rao (1946, 1947a,b,1948,1949a,b,1950) ([49-55]). He mainly followed the approach of Wald. The importance of these works of Rao is obvious. In addition to refining and generalizing the D^2-statistic, he suggested a measure of distance between two populations, discussed the problem of doubtful regions where definite decisions cannot be made and the generalization of the classification problem to three or more groups. Rao's development is for the case when the distributions are all known. General theoretical results on the classification problem in the frame work of decision theory are given in the book by Wald (1950, [67]) and in a paper by Wald and Wolfowitz (1950, [68]).

In what follows we shall be concerned only with the significant development in the theory of classification problems associated with two multivariate normal populations with unequal covariance matrices. Kullback (1952,1959, [30], [31]) considered a rule based on a linear statistic which maximizes the

Kullback–Leibler distance between two univariate normal
distributions of a linear statistic under the two hypotheses.
He also obtained some partial results on deriving the optimum
class of rules based on linear functions of observations from
Neyman–Pearson view-point (i.e. minimizing one probability of
misclassification by controlling the other). Clunies–Ross
and Riffenburgh (1960, [16]) studied this problem geometrically.
Anderson and Bahadur (1962, [4]) derived the minimax rule and
characterized the minimal complete class after restricting to
the class of rules based on linear functions of observations.

The distribution of the quadratic discriminant function
was studied by Okamoto (1963, [42]) for the special case
$\underset{\sim}{\mu}_1 = \underset{\sim}{\mu}_2$; by Bartlett and Please (1963, [9]) for the special
case $\underset{\sim}{\mu}_1 = \underset{\sim}{\mu}_2 = 0$ and

$$R_j = \begin{bmatrix} 1 & b_j & b_j & \cdots & b_j \\ b_j & 1 & b_j & \cdots & b_j \\ \vdots & & & & \vdots \\ b_j & b_j & \cdots\cdots\cdots & 1 \end{bmatrix}, \quad (j = 1,2).$$

Matusita (1967, [38]) suggested a minimum distance rule
(MDR). Let X be the random variable under consideration, and
S_n the empirical distribution based on n observations on X.
Suppose that X has one of F_1, F_2 as its distribution. Let
S'_r, S''_s be the empirical distributions determined by observations

on F_1 and F_2 respectively. Then the decision rule he proposed
is the following :

(i)   when $d(S_n,S_r') < d(S_n,S_s'')$, we decide on F_1,

(ii) when $d(S_n,S_r') > d(S_n,S_s'')$, we decide on F_2,

(iii) For the case $d(S_n,S_r') = d(S_n,S_s'')$, we determine in
advance to take either of F_1,F_2 , say F_1.

Here, $d(.,.)$ is a distance in the space of distributions
concerned. He took

$$d(F_1,F_2) = \left[\int_\Omega \left(\sqrt{p_1}(x) - \sqrt{p_2}(x)\right)^2 dm\right]^{\frac{1}{2}},$$

where F_j are defined on Ω , with p.d.f. p_j with respect to
a σ-finite measure m. He studied seperately the cases according
as the $\underset{\sim}{\mu}_j$'s and R_j's are known or unknown , and obtained some
bounds on the probability of correct classification (using MDR)
and on the total probability of misclassification resulting
from the use of Bayes rule.

When $R_1 = dR_2$ $(d > 1)$, the distributions of the QDF and
its plug in version (i.e. by replacing the parameters by their
estimates) were studied by Han (1969, [23]) . Similar results
were obtained by Han (1970, [24]) when the R_j's are of
"circular" type. The same problem was treated by Gilbert
(1969, [19]). The author compared the total probability of
misclassification (PMC) resulting from the use of QDF with
that of the LDF:

$$\delta'(\omega R_1 + \overline{1-\omega}R_2)^{-1}x \ ,$$

where $\delta = \mu_1 - \mu_2$, ω = prior probability of H_1 assuming parameters are known. For the latter the optimum cut-off point for which the total PMC is minimized was obtained.

Shumway and Unger (1974, [59]) considered LDFs for stationary time series. They obtained the asymptotic optimal LDF corresponding to the criterion of maximizing Kullback-Leibler distance using certain spectral approximations. The optimizing discriminant function was applied to the seismic data from selected earthquakes and nuclear explosions.

The recent book by Seber (1984, [58]) contains interesting discussions on discriminant analysis methodology and , in addition, includes an extensive bibliography.

1.3 SUMMARY

The thesis is divided into seven chapters. Chapter II provides some definitions and elementary results needed in subsequent chapters of the present work.

Chapter III contains a systematic study of optimal classfication rules based on linear statistics which maximize the Bhattacharyya distance (B-D) in the case when the observed process is discrete in time. Both stationary and non-stationary cases have been considered. For an arbitrary time series, we have to solve iteratively an implicit equation in order to

get our desired linear discriminant function (LDF); we have given a simple method of ascertaining an interval of convergence of the iteration process. Linear procedures based on the B-D belong to the Anderson-Bahadur admissible class for a proper choice of the cut-off point. Some special cateogories of problems where the mean vectors and the covariance matrices are of specific kind have also been studied here. We compare the performance of our LDF with that of some other LDFs considered in the literature. The comparison with the quadratic discriminant function due to the Bayes' criterion is considered when covariance matrices are proportional. These comparisons result in the conclusion that the distance of our interest is worth-considering. A one-to-one correspondence between the two classes of linear procedures —— one due to our criterion of maximizing the B-D and another of minimizing the total probability of misclassification subject to a linear relationship between the two types of errors —— has been established. Under certain regularity conditions, a compact form of the LDF for a covariance stationary time series when the sample size is large is obtained. Some illustrative examples satisfying the regularity conditions are given and it is shown that errors of misclassification tend to zero asymptotically.

Chapter IV deals with the continous time series. It is shown that finding the optimal LDF amounts to solving an integral equation of Fredholm type. We are able to obtain a compact form of the LDF in the case when the time series is covariance stationary with the observation interval infinite.

Chapter V is devoted to designing of signals. It is shown that even in the simple case when the two types of errors are made equal we are unable to obtain an explicit expression for the optimum signal. An analytical solution for the optimal signal however is available only through a bound on the total probability of misclassification.

Since complex normal processes are of interest in many applied areas ; in Chapter VI all the major results of the preceeding chapters have been extended to the case where the underlying process is complex valued. Here too, some special categories of problems of practical interest have been discussed.

The concluding Chapter VII includes some suggestions for further investigation in this interesting field.

Throughout this work, the underlying process is assumed to be real valued, unless stated otherwise (Chapter VI).

CHAPTER II

PRELIMINARIES

2.1 INTRODUCTION

In this chapter, some definitions and elementary results are collected for later use in the present work. The basic idea of statistical distance is given in Section 2.2. We define in Section 2.3 the Bhattacharyya distance between two populations which is of primary concern to our investigation and for easy reference, we derive the Bhattacharyya distance between two normal populations. This is followed by a remark where it is shown how this distance measures the dissimilarity between them. The relevent notions of covariance stationarity and spectral density of a stochastic process are contained in Section 2.4 while Section 2.5 contains some basic concepts of complex normal processes.

2.2 STATISTICAL DISTANCE MEASURES

The concept of distance associated with a metric space is well known. If x and y are two "points" in a metric space, then $d(x,y)$, called the distance between x and y, satisfies the three properties :

    i)   non-negativity

    ii) symmetry

    iii) triangle inequality

As an example take the metric space $\mathbb{R}^n$ and $d(\underset{\sim}{x},\underset{\sim}{y}) = \{\sum_{i=1}^{n} |x_i - y_i|^p\}^{\frac{1}{p}}$, $p \geq 1$, for $\underset{\sim}{x}, \underset{\sim}{y} \in \mathbb{R}^n$. These various distance functions can be

viewed as measures of dissimilarity between two "points" in the metric space $\mathbb{R}^n$.

Analogously, several indices have been suggested in statistical literature to reflect the degree of dissimilarity between any two probability distributions. Such indices have been variously called measures of distance between two distributions (see [1], for instance), measures of seperation ([56]), measures of discriminatory information ([15,31]), and measures of variation-distance ([28]). While these indices have not all been introduced for exactly the same purpose, as the names given to them imply, they have the common property of increasing as the two distributions involved ''move apart''. An index with this property may be called a measure of divergence of one distributi from another. A general method of generating measures of divergence has been discussed in a paper of Ali and Silvey ([2]) and it is shown therein that various available measures of divergence belong to a general class defined by their method.

We shall now give some examples of measures of divergence. Let $(\Omega, \mathcal{B}, \nu)$ be a measure space and $\mathcal{P}$ be the set of all probability measures on $\mathcal{B}$ which are absolutely continuous with respect to ν. Consider two such probability measures $P_1, P_2 \varepsilon \mathcal{P}$ and let p_1 and p_2 be their respective density functions with respect to ν. Then

(i) $\quad J(1,2) = \int_{\Omega} (p_2 - p_1) \ln \frac{p_2}{p_1} d\nu$

is known as Jeffrey's measure of divergence

(ii) $I(1,2) = \int_{\Omega} p_1 \ln \frac{p_1}{p_2} d\nu$

and $I(2,1) = \int_{\Omega} p_2 \ln \frac{p_2}{p_1} d\nu$

are called Kullback-Leibler's measures of discriminatory information ([31]).

(iii) Kolmogorov's measure of variational distance ([28]) is given by

$$\frac{1}{2} \int_{\Omega} |\sqrt{p_2} - \sqrt{p_1}| \, d\nu$$

(iv) Matusita's measure of distance ([36]) is as follows :

$$\int_{\Omega} (\sqrt{p_2} - \sqrt{p_1})^2 \, d\nu$$

(v) Chernoff's ([15]) measure of discriminatory information is defined by

$$- \ln \inf_{0<t<1} (\int_{\Omega} p_1^t p_2^{t-1} \, d\nu)$$

Remark 2.2.1. Statistical distance functions need not satisfy all the properties of a distance function stated at the beginning of Section 2.2. We note in Section 2.3 that the Bhattacharyya distance may not satisfy the triangle inequality.

2.3 BHATTACHARYYA DISTANCE

In this work we shall be concerned with the Bhattacharyya distance which was first introduced in a statistical context by Bhattacharyya ([11]). An early and well known statistical

application of this measure was made by Kakutani ([27]) who himself mentions an earlier appearance of this measure in a non-statistical problem (see Hellinger [25]). Therefore, the names of Hellinger and Kakutani are also often associated with the Bhattacharyya distance.

We first define the Bhattacharyya coefficient for the densities p_i (i = 1,2) by

$$\rho_2(p_1,p_2) \triangleq \int_\Omega (p_1 p_2)^{\frac{1}{2}} d\nu \qquad (2.3.1)$$

More generally,

$$\rho_2(P_1,P_2) \triangleq \int_\Omega (\frac{dP_1}{d\nu} \frac{dP_2}{d\nu})^{\frac{1}{2}} d\nu$$

where $\frac{dP_i}{d\nu}$ is the Radon-Nikodym derivative of P_i (i = 1,2) with respect to ν. $\rho_2(P_1,P_2)$ is called the Bhattacharyya coefficient or affinity between the two probability measures P_1 and P_2. Note that $\rho_2(P_1,P_2)$ does not depend on the measure ν dominating P_1 and P_2. $\rho_r(P_1,\ldots,P_r)$, the affinity among r d.f.s. $P_1,\ldots,P_r$, can be defined (see [37]) analogously.

The following proposition states some properties of $\rho_2(P_1,P_2)$ which can be proved easily (see, for example [39]).

PROPOSITION 1 :

    i) $0 \leq \rho_2(P_1,P_2) \leq 1$

    ii) $\rho_2(P_1,P_2) = 1$ if and only if $P_1 = P_2$

    iii) $\rho_2(P_1,P_2) = 0$ if and only if $P_1 \perp P_2$.

The Bhattacharyya distance between two probability distributions P_1 and P_2 is defined by

$B \overset{\Delta}{=}$ the Bhattacharyya distance

$$= -\ln \rho_2(P_1, P_2) \tag{2.3.2}$$

Clearly, $0 \leqq B \leqq \infty$.

The Bhattacharyya distance need not satisfy the triangle inequality (see Appendix A).

We shall now derive the Bhattacharyya distance between two multivariate normal populations.

PROPOSITION 2 : Let $\prod_1$ and $\prod_2$ be two n-variate normal populations with distributions $N(\mu_1, R_1)$ and $N(\mu_2, R_2)$ respectively, where R_1 and R_2 are positive definite matrices. Then

$$-\ln \rho_2 = \frac{1}{4}\rho + \frac{1}{8}D^2 \tag{2.3.3}$$

where $\rho \overset{\Delta}{=} 2\ln |R| - \ln |R_1| - \ln |R_2|$,

$$D^2 \overset{\Delta}{=} (\mu_1 - \mu_2)' R^{-1}(\mu_1 - \mu_2)$$

and $R \overset{\Delta}{=} (R_1 + R_2)/2$.

Proof : Let the random vector X have the densities p_j under $\prod_j$ ($j = 1,2$) respectively. Then by definition,

$$\rho_2(1,2;X) = \int_\Omega (p_1(X)p_2(X))^{\frac{1}{2}} dX$$

(for convenience, we write $\rho_2(1,2;X)$ in place of $\rho_2(P_1, P_2;X)$)

$$= \left[(2\pi)^{2n} \, |R_1| |R_2| \right]^{-\frac{1}{4}} \int_{\mathbb{R}^n} \exp\left[-\frac{1}{4} \{ (x-\mu_1)' R_1^{-1} (x-\mu_1) \right.$$

$$\left. + (x-\mu_2)' R_2^{-1} (x-\mu_2) \} \right] dx$$

$$(2.3.4)$$

Using Appendix B , we can write

$$(x-\mu_1)' R_1^{-1} (x-\mu_1) + (x-\mu_2)' R_2^{-1} (x-\mu_2)$$

$$= (x-m)' \, S(x-m) + (\mu_1-\mu_2)' (R_1+R_2)^{-1} (\mu_1-\mu_2)$$

where $m = (R_1+R_2)^{-1}(R_2\mu_1 + R_1\mu_2)$

and $\quad S = R_1^{-1}(R_1+R_2)R_2^{-1}$.

These relations when used in (2.3.4) give

$$\rho_2(1,2;x) = \frac{(|R_1| |R_2|)^{\frac{1}{4}}}{|\frac{1}{2}(R_1+R_2)|^{\frac{1}{2}}} \; \exp\{ -\frac{1}{4}(\mu_1-\mu_2)'(R_1+R_2)^{-1}(\mu_1-\mu_2) \}$$

$$(2.3.5)$$

Hence the result.

<u>Remark 2.3.1</u> : Putting $n = 1$, we have

$$\rho_2(1,2;x) = \frac{(\eta_1^2 \eta_2^2)^{\frac{1}{4}}}{\{ \frac{1}{2}(\eta_1^2+\eta_2^2) \}^{\frac{1}{2}}} \exp\{ -\frac{1}{4}\frac{(m_1-m_2)^2}{\eta_1^2+\eta_2^2} \}$$

$$(2.3.6)$$

where the involved distributions are $N(m_1, \eta_1^2)$ and $N(m_2, \eta_2^2)$.

<u>Remark 2.3.2</u> : (1) The Bhattacharyya distance comes out as a special case of the Chernoff distance taking $t = \frac{1}{2}$.

(2) We note that an n-variate normal distribution is completely specified by the set of $\frac{n(n+3)}{2}$ parameters,

$$\{ \nu_1, \ldots, \nu_n \, , \, \lambda_{11}, \ldots, \lambda_{nn}, \lambda_{12}, \ldots, \lambda_{(n-1)n} \}$$

of which the first n are the location parameters and the
remaining are the orientation parameters. Naturally, the
dissimilarity between the populations can be judged through
the disagreement between the corresponding location and
orientation parameters of the two populations. Now, an
examination of the expression (2.3.3) reveals that the
first term measures the dissimilarity of the two populations
with respect to their orientations, while the second term
does so in terms of their locations. In other words, ρ measures
the divergence in the dispersion matrices and D^2 between the
mean values, and the total divergence is a weighted sum of the
two.

2.4 STOCHASTIC PROCESSES : SOME DEFINITIONS AND RESULTS

Let $\{X(t), t\epsilon T\}$ be a stochastic process defined on
a probability space $(\Omega, \mathcal{J}, P)$ where the index set T may be
discrete or continuous. In discrete case, T may be one
of the forms $\{o, \pm 1, \pm 2, \ldots\}$ and $\{o, 1, 2, \ldots\}$ and in
the continuous case T may be $\{t : t \geq 0\}$ or $\{t : -\infty < t < \infty\}$.
The mean function is defined by $\mu(t) \overset{\Delta}{=} EX(t)$, $t \epsilon T$. The function

$$K(t_1, t_2) = Cov(X(t_1), X(t_2)) = E(X(t_1) - u(t_1))(X(t_2) - u(t_2)),$$

for $t_1, t_2 \epsilon T$, is called the covariance kernel of the process.

A stochastic process is said to be covariance stationary ([44])
it possesses finite second moments and if its covariance kernel

$K(t,s)$ is a function only of the absolute difference $|s-t|$, in the sense that there exists a function $R(v)$ such that for all s and t in T,

$$K(t,s) = R(|s-t|) \ ,$$

or more precisely, $R(v)$ has the property that for every t and $v \ \varepsilon \ T$,

$$Cov(X(t),X(t+v)) = R(v) \tag{2.4.1}$$

We call $R(v)$ the <u>covariance function</u> of the covariance stationary stochastic process $\{X(t), t\varepsilon T\}$.

A stochastic process with discrete time parameter which is covariance stationary defines a sequence of covariances, say, $r(o),r(1),\ldots$. The Fourier transform of the sequence is defined by (see [5]),

$$f(\lambda) = \sum_{n=-\infty}^{\infty} r(n) e^{j\lambda n} \ , \ -\pi \leq \lambda \leq \pi \tag{2.4.2}$$

provided $\sum_{n=-\infty}^{\infty} |r(n)| < \infty$, which will ensure that the series (2.4.2) converges uniformly and absolutely to the function $f(\lambda)$ (see [18]). The function $f(\lambda)$ is called the <u>spectral density</u> associated with $\{r(n)\}_{n=o}^{\infty}$ or of the process $\{X(t), t\varepsilon T\}$.

From (2.4.2), one can obtain,

$$r(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-i\lambda n} f(\lambda) \ d\lambda \tag{2.4.3}$$

For continuous time-parameter covariance stationary process the spectral density is analogously defined as

$$f(\lambda) = \int_{-\infty}^{\infty} r(h) e^{i\lambda h} \ dh \tag{2.4.4}$$

provided $\int_{-\infty}^{\infty} |r(v)| dv < \infty$, and then $r(v)$ is given by

$$r(v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda v} f(\lambda) d\lambda \qquad (2.4.5)$$

Since $r(n) = r(-n)$ ∀ n , (2.4.2) and (2.4.3) can be rewritten as

$$f(\lambda) = \sum_{n=-\infty}^{\infty} r(n) \cos \lambda n \qquad (2.4.6)$$

and $$r(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos \lambda n \, f(\lambda) d\lambda \qquad (2.4.7)$$

2.5 COMPLEX NORMAL PROCESSES

A complex stochastic process $\{ Z(t), t\varepsilon T \}$ is said to be a complex normal process if the real vector $(\text{Re } \underset{\sim}{z}, \text{Im } \underset{\sim}{z})'$ of $2n$ dimension has a $2n$-variate multivariate normal distribution with mean $(\text{Re } \underset{\sim}{\mu}, \text{Im } \underset{\sim}{\mu})'$ $\qquad\qquad (2.5.1)$

and covaria nce matrix

$$\frac{1}{2} \begin{bmatrix} \text{Re } R & -\text{Im } R \\ \text{Im } R & \text{Re } R \end{bmatrix}$$

where $\quad \underset{\sim}{z} = (Z(t_1), \ldots, Z(t_n))'_{nx1}$

$\underset{\sim}{\mu} \overset{\Delta}{=} E \underset{\sim}{z}$

$R \overset{\Delta}{=} E(\underset{\sim}{z} - \underset{\sim}{\mu})(\underset{\sim}{\bar{z}} - \underset{\sim}{\bar{\mu}})'$

for any choice of distinct points $t_1, t_2, \ldots, t_n$ in T and for any n. R is assumed to be a positive definite Hermitian matrix. $\underset{\sim}{\bar{z}}$ denotes the complex conjugate of $\underset{\sim}{z}$.

A complex random vector $\underset{\sim}{z} = \underset{\sim}{x} + i \underset{\sim}{\chi}$ is said to have n-dimensional complex multivariate normal distribution with mean $\underset{\sim}{\mu}$

and covariance matrix R if $(\underset{\sim}{x},\underset{\sim}{y})'$ of 2n dimension is distributed as a multivariate normal vector with mean vector and covariance matrix as specified in (2.5.1) and (2.5.2) respectively.

The density function of $\underset{\sim}{z}$ is given by

$$p(\underset{\sim}{z}) = \frac{1}{\pi^n \, |R|} \exp\{-(\bar{\underset{\sim}{z}}-\bar{\underset{\sim}{\mu}})' R^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu})\} \tag{2.5.3}$$

which is a real-valued scalar function of the complex vector $\underset{\sim}{z}$.

We note $p(\underset{\sim}{z}) \geq 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(\underset{\sim}{z}) \, d\underset{\sim}{x} \, d\underset{\sim}{y} = 1.$
One can easily show :

$$E(\underset{\sim}{z}-\underset{\sim}{\mu})(\underset{\sim}{z}-\underset{\sim}{\mu})' = 0 \tag{2.5.4}$$

We note that if $\underset{\sim}{z} = \underset{\sim}{x} + i\underset{\sim}{y}$ has n-dimensional complex multivariate normal distribution with mean vector $\underset{\sim}{\mu}$ and positive definite covariance matrix R , then the marginal density of $\underset{\sim}{x}$ is n-dimensional normal with mean Re $\underset{\sim}{\mu}$ and covariance matrix $\frac{1}{2}$ Re R. Similarly, $\underset{\sim}{y}$ has a n-variate normal distribution with mean Im $\underset{\sim}{\mu}$ and the same covariance matrix $\frac{1}{2}$ Re R.

In the following we define the spectral density of a covarianc stationary time series :

$W(\lambda) = \sum_{h=-\infty}^{\infty} \sigma(h) \, e^{-i\lambda h}$ is called the spectral density of of the time series $\{Z(t), t\varepsilon T\}$.

$\sigma(h) = Cov(Z(t), \overline{Z(t+h)})$, $-\pi \leq \lambda \leq \pi,$

provided $\sum_{-\infty}^{\infty} |\sigma(h)| < \infty.$

All the material presented in this section can be found in Miller ([40]).

CHAPTER III

LINEAR DISCRIMINANT FUNCTIONS FOR DISCRETE TIME SERIES

3.1 INTRODUCTION

In this chapter, we consider observations which constitute a discrete time series. The case of continuous time series is taken up in the next chapter. In Section 3.2, the problem of two-group classification is mathematically formulated as one of finding an optimal classification rule based on linear statistics which maximize the Bhattacharyya distance associated with the groups involved. We obtain in Section 3.3 an expression for the optimal linear discriminant function (LDF) in the sense of maximizing the Bhattacharyya distance and show that the resulting linear procedure belongs to the Anderson-Bahadur admissible class for a proper choice of the cut-off point. In the same section, a simple method of ascertaining the interval of convergence of the iteration process involved in finding the LDF is discussed. Some special categories of problems are also studied here. In this section we compare the performance of our LDF with that of some other LDFs considered in the literature. The comparison with the quadratic discriminant function due to the Bayes' criterion is presented in Section 3.4 when covariance matrices are proportional. A one-to-one correspondence between the two classes of linear procedures----

one due to our criterion of maximizing the Bhattacharyya
distance and another of minimizing the total probability
of misclassification subject to a linear relationship
between the two types of errors ---- has been established
in Section 6.5. Section 6.6 provides , under certain
regularity conditions, a compact form of the LDF for
covariance stationary time series when the sample size
is large. In the same section some illustrative examples
are given and it is shown that errors of misclassification
tend to zero asymptotically.

3.2 MATHEMATICAL FORMULATION OF THE PROBLEM

The problem is to classify a normal time series
$X = (X(o),..., X(n-1))$ as belonging to one of the two
populations described by the two hypotheses H_1 and H_2.
These hypotheses specify that the nx1 normal time series
X has means and covariances μ_1, R_1 and μ_2, R_2 under H_1 and
H_2 respectively.

A typical problem of this kind arises in communication
engineering. We shall consider this example in Chapter V
from the view-point of ''signal design''.

The admissible procedure for classification provided
by the Neyman-Pearson theory as well as the Bayes' rule are
based on the likelihood ratio. If $R_1 = R_2$, the likelihood
ratio depends on a linear function of X (called the Fisher's

discriminant function); this case has been extensively
studied in the literature ([3]). If $R_1 \neq R_2$, the Bayes'
rule states :

assign $\underset{\sim}{x}$ to H_1 (or H_2) according as

$$[\frac{1}{2} \ln \frac{|R_2|}{|R_1|} - \frac{1}{2} \underset{\sim}{\mu_1'} R_1^{-1} \underset{\sim}{\mu_1} + \frac{1}{2} \underset{\sim}{\mu_2'} R_2^{-1} \underset{\sim}{\mu_2}$$

$$- \frac{1}{2} \{\underset{\sim}{x'} (R_1^{-1} - R_2^{-1})\underset{\sim}{x} - 2\underset{\sim}{x'} (R_1^{-1}\underset{\sim}{\mu_1} - R_2^{-1}\underset{\sim}{\mu_2})\}] > k(<k)$$

$$(3.2.1)$$

$$k \overset{\Delta}{=} \ln(\omega_2 \, C(1/2)/\omega_1 \, C(2/1))$$

where ω_j is the prior probability of H_j and $C(j/i)$ is
the cost of misclassifying an $\underset{\sim}{x}$ to H_j when $\underset{\sim}{x}$ actually
belongs to H_i ($i \neq j, i, j = 1, 2$).(Note that $C(1/1) = C(2/2) = 0$).
The quantity $\underset{\sim}{x'} (R_1^{-1} - R_2^{-1})\underset{\sim}{x} - 2\underset{\sim}{x'} (R_1^{-1}\underset{\sim}{\mu_1} - R_2^{-1}\underset{\sim}{\mu_2})$ is called
the quadratic discriminant function (QDF), and in the case
of unequal covariance matrices, one has to use a QDF since
$(R_1^{-1} - R_2^{-1})$ does not vanish. Unfortunately, the distribution
theory pertaining to the quadratic part involves weighted
sum of non-central chi-square random variables ([59]) so
that computing error rates resulting from its use seems
difficult. Hence the usual approach has been ([4,31,59])
to consider a linear procedure ; in other words, in a situation
where the two populations are centered around different points
and have different patterns of scatter, and where one considers
multivariate normal distributions to be reasonably good
approximations for those two populations, one may want to

divide the sample space into two regions of classification by some simple curve or surface. The simplest is a line or hyperplane ; the procedure may then be termed linear.

We now <u>define</u> linear procedures formally. Let $\underset{\sim}{\alpha} \neq o$ be a nx1 vector and c be a scalar. An observation $\underset{\sim}{x}$ is classified as coming from the population under H_1 if $\underset{\sim}{\alpha}'\underset{\sim}{x} \geq c$ and from H_2 otherwise. Briefly, we write :

$$\underset{\sim}{\alpha}'\underset{\sim}{x} \underset{H_2}{\overset{H_1}{\gtrless}} c \qquad\qquad (3.2.2)$$

(accept)

Thus the problem is to find $\underset{\sim}{\alpha}$ and c in some optimal way. We assume $\underset{\sim}{\mu}_1 \neq \underset{\sim}{\mu}_2$ and R_1 and R_2 (possibly unequal) to be positive definite covariance matrices. The parameters are assumed to be known. Under these conditions, we study the classification rule based on linear statistics which is optimal with respect to our criterion of maximizing the Bhattacharyya distance. We call $\underset{\sim}{\alpha}'\underset{\sim}{x}$ the ''linear discriminant function''([4]).

3.3 METHOD FOR OBTAINING $\underset{\sim}{\alpha}$ IN THE DISCRIMINANT FUNCTION

Under H_1 and H_2 , the parameters of the normal distribution of the linear discriminant function $y = \underset{\sim}{\alpha}'\underset{\sim}{x}$ are :

$$E_{H_1}(y) = \underset{\sim}{\alpha}'\underset{\sim}{\mu}_1, \ E_{H_2}(y) = \underset{\sim}{\alpha}'\underset{\sim}{\mu}_2, \ Var_{H_1}(y) = \underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}, \ Var_{H_2}(y) = \underset{\sim}{\alpha}'R_{}$$

$$(3.31)$$

It follows from (2.2.3) that

$$\ln \rho_2(1,2;y) = \frac{1}{4}[\ln(\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}) + \ln(\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha})] - \frac{1}{2}\ln\frac{1}{2}\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}$$

$$- \frac{1}{4}\frac{(\underset{\sim}{\alpha}'\underset{\sim}{\delta})^2}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}} \qquad (3.3.2)$$

where $\qquad \underset{\sim}{\delta} \overset{\Delta}{=} \underset{\sim}{\mu_1} - \underset{\sim}{\mu_2}.$

Differentiating ([57]) (3.3.2) with respect to $\underset{\sim}{\alpha}$, we get

$$\frac{\partial \ln\rho_2}{\partial\underset{\sim}{\alpha}} = \frac{1}{4}\left[\frac{2R_1\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}} + \frac{2R_2\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha}}\right] - \frac{1}{2}\frac{2\frac{1}{2}(R_1+R_2)\underset{\sim}{\alpha}}{\frac{1}{2}\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}$$

$$- \frac{1}{4}\left[\frac{2(\underset{\sim}{\alpha}'\underset{\sim}{\delta})\underset{\sim}{\delta}\{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\} - (\underset{\sim}{\alpha}'\underset{\sim}{\delta})^2 \cdot 2(R_1+R_2)\underset{\sim}{\alpha}}{\{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\}^2}\right]$$

$$= \frac{1}{2}\left[\frac{R_1\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}} + \frac{R_2\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha}}\right] - \frac{(R_1+R_2)\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}$$

$$-\frac{1}{2}\left[\frac{(\underset{\sim}{\alpha}'\underset{\sim}{\delta})\underset{\sim}{\delta}\{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\} - (\underset{\sim}{\alpha}'\underset{\sim}{\delta})^2\{(R_1+R_2)\underset{\sim}{\alpha}\}}{\{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\}^2}\right]$$

Set $\dfrac{\partial \ln\rho_2}{\partial\underset{\sim}{\alpha}} = o$

$$\Rightarrow \frac{(\underset{\sim}{\alpha}'\underset{\sim}{\delta})\underset{\sim}{\delta}}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}} = \frac{(\underset{\sim}{\alpha}'\underset{\sim}{\delta})^2(R_1+R_2)\underset{\sim}{\alpha}}{\{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\}^2} + \frac{R_1\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}} + \frac{R_2\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha}} - \frac{2(R_1+R_2)\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}$$

$$= \left[\frac{(\underset{\sim}{\alpha}'\underset{\sim}{\delta})^2}{\{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\}^2} + \frac{1}{\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}} - \frac{2}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}\right]R_1\underset{\sim}{\alpha}$$

$$+ \left[\frac{(\underset{\sim}{\alpha}'\underset{\sim}{\delta})^2}{\{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\}^2} + \frac{1}{\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha}} - \frac{2}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}\right] R_2\underset{\sim}{\alpha}$$

$\Rightarrow (t_1 R_2 + t_2 R_2)\underset{\sim}{\alpha} = \underset{\sim}{\delta}$, where t_1, t_2 are given by

$$t_1 = \left[\frac{(\underset{\sim}{\alpha}'\underset{\sim}{\delta})^2}{\{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\}^2} + \frac{1}{\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}} - \frac{2}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}\right]/\left[\frac{\underset{\sim}{\alpha}'\underset{\sim}{\delta}}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}\right]$$

$$t_2 = \left[\frac{(\underset{\sim}{\alpha}'\underset{\sim}{\delta})^2}{\{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\}^2} + \frac{1}{\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha}} - \frac{2}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}\right]/\left[\frac{\underset{\sim}{\alpha}'\underset{\sim}{\delta}}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}\right]$$

$\Rightarrow \underset{\sim}{\alpha} = (t_1 R_1 + t_2 R_2)^{-1}\underset{\sim}{\delta}$, provided $(t_1 R_1 + t_2 R_2)$ is non-singular.

Therefore $\underset{\sim}{\alpha} = \frac{1}{t_1} R_\theta^{-1}\underset{\sim}{\delta}$, (3.3.3)

where $R_\theta \overset{\Delta}{=} R_1 - \theta R_2$ (3.3.4)

and $-\theta \overset{\Delta}{=} \frac{t_2}{t_1}$ (3.3.5)

That is, the value of $\underset{\sim}{\alpha}$ for which $-\ln\rho_2(1,2;y)$ is a maximum satisfies (3.3.3).

<u>Remark 3.3.1</u> : $-\ln\rho_2(1,2;\underset{\sim}{\alpha}'\underset{\sim}{x})$ is invariant under scalar multiplication of $\underset{\sim}{\alpha}$. It follows immediately, if we write (3.3.2) in the following form :

$$\ln\rho_2(1,2;\underset{\sim}{\alpha}'\underset{\sim}{x}) = \ln\frac{\{(\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha})(\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha})\}^{1/4}}{\{\frac{1}{2}\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\}^{1/2}} - \frac{1}{4}\frac{(\underset{\sim}{\alpha}'\underset{\sim}{\delta})^2}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}$$

(3.3.2)'

Remark 3.3.2 : If $\underset{\sim}{\alpha}_*$ is a solution of (3.3.3), so is $\underset{\sim}{\alpha}_{**} = k\underset{\sim}{\alpha}_*$, k being any non-zero scalar. It follows trivially from (3.3.3).

Remark 3.3.3 : Following the above remarks, we can say that the maximization of $-\ln\rho_2(1,2;\underset{\sim}{\alpha}'\underset{\sim}{x})$ is irrespective of the value of $1/t_1$ attached as a factor to $R_\theta^{-1}\underset{\sim}{\delta}$. Hence $\underset{\sim}{\alpha} = \frac{1}{\tau} R_\theta^{-1}\underset{\sim}{\delta}$ and $\underset{\sim}{\alpha} = \frac{1}{t_1} R_\theta^{-1}\underset{\sim}{\delta}$, where τ is any non-zero scalar, gives the same optimal solutions. Thus the problem of determining t_1 and t_2 reduces to finding of the ratio t_2/t_1. We can assign any value to τ ; this would not affect the desired maximization process. We take $\tau = 1$.

Consequently, the required optimal solutions are necessarily of the form :

$$\underset{\sim}{\alpha} = R_\theta^{-1}\underset{\sim}{\delta} \qquad (3.3.6)$$

where

$$-\theta = \frac{\{\dfrac{\underset{\sim}{\alpha}'\underset{\sim}{\delta}}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}\}^2 + \dfrac{1}{\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha}} - \dfrac{2}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}}{\{\dfrac{\underset{\sim}{\alpha}'\underset{\sim}{\delta}}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}\}^2 + \dfrac{1}{\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}} - \dfrac{2}{\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}}} \qquad (3.3.7)$$

It is clear that (3.3.6) is an implicit equation in $\underset{\sim}{\alpha}$. Hence an iterative procedure must be employed to solve for $\underset{\sim}{\alpha}$.

3.3.1 AN EXAMPLE

Initial or entering values of $\underset{\sim}{\alpha}$ are required to begin an iterative procedure. The entering value for $\underset{\sim}{\alpha}$ is taken as

$$\underset{\sim}{\alpha}^{(o)} = (R_1 + R_2)^{-1} \underset{\sim}{\delta}$$

With $\underset{\sim}{\alpha}^{(o)}$ determined, values for $\underset{\sim}{\alpha}^{(o)'}\underset{\sim}{\delta}, \underset{\sim}{\alpha}^{(o)'}R_1\underset{\sim}{\alpha}^{(o)}, \underset{\sim}{\alpha}^{(o)'}R_2\underset{\sim}{\alpha}^{(o)}$ are found and then $\theta^{(o)}$. Cycle 1 is begun by entering with $\theta^{(o)}$ to find a new $\underset{\sim}{\alpha}$ from

$$\underset{\sim}{\alpha}^{(1)} = R^{-1}_{\theta^{(o)}}\underset{\sim}{\delta}$$

and then determining $\underset{\sim}{\alpha}^{(1)'}\underset{\sim}{\delta}, \underset{\sim}{\alpha}^{(1)'}R_1 \underset{\sim}{\alpha}^{(1)}, \underset{\sim}{\alpha}^{(1)'}R_2\underset{\sim}{\alpha}^{(1)}$ and then $\theta^{(1)}$, thus completing the first cycle. This procedure is continued until the difference in successive θ's is as small as desired.

We shall illustrate the procedure described with data given in Kullback (Chapter 13, [31]).

Example 3.3.1

$$\underset{\sim}{X} \sim N_2 \left(\begin{pmatrix} 20.80 \\ 12.32 \end{pmatrix}, \begin{pmatrix} 6.92 & -5.27 \\ -5.27 & 40.89 \end{pmatrix} \right) \text{ under } H_1$$

$$\underset{\sim}{X} \sim N_2 \left(\begin{pmatrix} 12.80 \\ 36.40 \end{pmatrix}, \begin{pmatrix} 36.75 & 13.92 \\ 13.92 & 287.92 \end{pmatrix} \right) \text{ under } H_2.$$

Then , $\underset{\sim}{\delta} = \underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2 = \begin{pmatrix} 8.00 \\ -24.08 \end{pmatrix}$.

We have programmed the procedure (see Appendix C). The result correct up to 4 decimal places is as follows (number of iterations taken is 3) :

$$\underset{\sim}{\alpha}_* = (1,-.4153)'.$$

Thus the linear discriminant function is : $y = x_1 - 0.4153 \quad x_2$.
The value of θ for which we get $\underset{\sim}{\alpha}_*$ is given by $\theta_* = -.4152$.

3.3.2 CONVERGENCE OF THE ITERATION PROCESS

Since R_1 and R_2 are positive definite matrices by assumption, there always exists a non-singular matrix P such that

$$R_1 = P'P$$

and
$$R_2 = P'\Lambda P = P' \begin{bmatrix} \lambda_1 & & O \\ & \ddots & \\ O & & \lambda_n \end{bmatrix} P ,$$

where λ_i's $(i = 1,n)$ are the characteristic roots of $R_2 R_1^{-1}$ (see [57]).

Then $\underset{\sim}{\alpha}' R_1 \underset{\sim}{\alpha} = \underset{\sim}{\alpha}' P'P \underset{\sim}{\alpha} = \underset{\sim}{\beta}'\underset{\sim}{\beta}$, say, where $P\underset{\sim}{\alpha} = \underset{\sim}{\beta}$,

$$\underset{\sim}{\alpha}' R_2 \underset{\sim}{\alpha} = \underset{\sim}{\alpha}' P'\Lambda P \underset{\sim}{\alpha} = \underset{\sim}{\beta}'\Lambda \underset{\sim}{\beta}$$

$$\underset{\sim}{\alpha}'\underset{\sim}{\delta} = \underset{\sim}{\beta}' (P')^{-1}\underset{\sim}{\delta} = \underset{\sim}{\beta}'\underset{\sim}{\eta} , \text{ say } , \text{ where } \underset{\sim}{\delta} = P'\underset{\sim}{\eta}$$

Now $\underset{\sim}{\beta} = P\underset{\sim}{\alpha} = P(R_1 - \theta R_2)^{-1} \underset{\sim}{\delta}$

$$= P[P'P - \theta P'\Lambda P]^{-1}\underset{\sim}{\delta}$$

$$= P[P'(I - \theta\Lambda)P]^{-1}\underset{\sim}{\delta}$$

$$= (I - \theta\Lambda)^{-1}(P')^{-1}\underset{\sim}{\delta}$$

$$= (I - \theta\Lambda)^{-1}\underset{\sim}{\eta}$$

i.e. $\underset{\sim}{\beta} = (I - \theta\Lambda)^{-1}\underset{\sim}{\eta}$.

Thus (3.3.7) reduces to

$$- \theta = \frac{\{\frac{\beta'\eta}{\beta'(\Lambda+I)\beta}\}^2 + \frac{1}{\beta'\Lambda\beta} - \frac{2}{\beta'(\Lambda+I)\beta}}{\{\frac{\beta'\eta}{\beta'(\Lambda+I)\beta}\}^2 + \frac{1}{\beta'\beta} - \frac{2}{\beta'(\Lambda+I)\beta}}$$

$$= \frac{\frac{2}{\beta'(\Lambda+I)\beta} - \frac{1}{\beta'\Lambda\beta} - \{\frac{\beta'\eta}{\beta'(\Lambda+I)\beta}\}^2}{\frac{2}{\beta'(\Lambda+I)\beta} - \frac{1}{\beta'\beta} - \{\frac{\beta'\eta}{\beta'(\Lambda+I)\beta}\}^2}$$

$$= \frac{A - \frac{1}{\beta'\Lambda\beta}}{A - \frac{1}{\beta'\beta}} \;, \qquad\qquad (3.3.8)$$

(where

$$A \triangleq \frac{2}{\beta'(\Lambda+I)\beta} - \{\frac{\beta'\eta}{\beta'(\Lambda+I)\beta}\}^2 \;)$$

$$= -\frac{A\theta}{A} \;, \text{ provided } A \neq 0 \;;$$

the case corresponding to A = o can be treated seperately.

Thus

$$- \theta = \frac{A+A\theta - \frac{1}{\beta'\Lambda\beta}}{-\frac{1}{\beta'\beta}}$$

$$= \frac{\beta'\beta}{\beta'\Lambda\beta} - A(1+\theta)\beta'\beta$$

or, $\theta = - \dfrac{\underset{\sim}{\beta}'\underset{\sim}{\beta}}{\underset{\sim}{\beta}'\Lambda\underset{\sim}{\beta}} + (1+\theta)(\dfrac{2}{\underset{\sim}{\beta}'(\Lambda+I)\underset{\sim}{\beta}} - \{\dfrac{\underset{\sim}{\beta}'\underset{\sim}{\eta}}{\underset{\sim}{\beta}'(\Lambda+I)\underset{\sim}{\beta}}\}^2)(\underset{\sim}{\beta}'\underset{\sim}{\beta})$

$\overset{\Delta}{=} \Psi(\theta)$

$$(3.3.9)$$

Taking the derivative of $\Psi(\theta)$ with respect to θ, we have,

$\Psi'(\theta) = - \dfrac{1}{(\underset{\sim}{\beta}'\Lambda\underset{\sim}{\beta})^2} \{(\sum\limits_{i=1}^{n} \dfrac{2\lambda_i\eta_i^2}{(1-\theta\lambda_i)^3})(\underset{\sim}{\beta}'\Lambda\underset{\sim}{\beta}) - (\underset{\sim}{\beta}'\underset{\sim}{\beta}) 2 \sum\limits_{i=1}^{n} \dfrac{\lambda_i^2\eta_i^2}{(1-\theta\lambda_i)^3}\}$

$\quad + \{\dfrac{2}{\underset{\sim}{\beta}'(\Lambda+I)\underset{\sim}{\beta}} - (\dfrac{\underset{\sim}{\beta}'\underset{\sim}{\eta}}{\underset{\sim}{\beta}'(\Lambda+I)\underset{\sim}{\beta}})^2 \} (\underset{\sim}{\beta}'\underset{\sim}{\beta})$

$\quad + 2(1+\theta)(\sum\limits_{i=1}^{n} \dfrac{\lambda_i\eta_i^2}{(1-\theta\lambda_i)^3}) \{ \dfrac{2}{\underset{\sim}{\beta}'(\Lambda+I)\underset{\sim}{\beta}} - (\dfrac{\underset{\sim}{\beta}'\underset{\sim}{\eta}}{\underset{\sim}{\beta}'(\Lambda+I)\underset{\sim}{\beta}})^2\}$

$\quad + (1+\theta)(\underset{\sim}{\beta}'\underset{\sim}{\beta})[- \dfrac{2}{\{\underset{\sim}{\beta}'(\Lambda+I)\underset{\sim}{\beta}\}^2}(\sum\limits_{i=1}^{n} \dfrac{2(1+\lambda_i)\lambda_i\eta_i^2}{(1-\theta\lambda_i)^3})$

$\quad - \dfrac{2\underset{\sim}{\beta}'\underset{\sim}{\eta}}{\{\underset{\sim}{\beta}'(\Lambda+I)\underset{\sim}{\beta}\}^3} \{(\underset{\sim}{\beta}'(\Lambda+I)\underset{\sim}{\beta})(\sum\limits_{i=1}^{n} \dfrac{\lambda_i\eta_i^2}{(1-\theta\lambda_i)^2})$

$\qquad\qquad - (\underset{\sim}{\beta}'\underset{\sim}{\eta})2\sum\limits_{i=1}^{n} \dfrac{\lambda_i(1+\lambda_i)\eta_i^2}{(1-\theta\lambda_i)^3}\}]$

$= -2 \dfrac{\sum\limits_{i=1}^{n} \dfrac{\lambda_i\eta_i^2}{(1-\theta\lambda_i)^3}}{\sum\limits_{i=1}^{n} \dfrac{\lambda_i\eta_i^2}{(1-\theta\lambda_i)^2}} + 2 \dfrac{\sum\limits_{i=1}^{n} \dfrac{\eta_i^2}{(1-\theta\lambda_i)^2}}{\{\sum\limits_{i=1}^{n} \dfrac{\lambda_i\eta_i^2}{(1-\theta\lambda_i)^2}\}^2}(\sum\limits_{i=1}^{n} \dfrac{\lambda_i^2\eta_i^2}{(1-\theta\lambda_i)^3})$

$$+ 2 \frac{\sum\limits_{i=1}^{n} \frac{\eta_i^2}{(1-\theta\lambda_i)^2}}{(\sum\limits_{i=1}^{n} \frac{(1+\lambda_i)\eta_i^2}{(1-\theta\lambda_i)^2})} - \frac{(\sum\limits_{i=1}^{n} \frac{\eta_i^2}{(1-\theta\lambda_i)})^2 (\sum\limits_{i=1}^{n} \frac{\eta_i^2}{(1-\theta\lambda_i)^2})}{(\sum\limits_{i=1}^{n} \frac{(1+\lambda_i)\eta_i^2}{(1-\theta\lambda_i)^2})^2}$$

$$+ 4(1+\theta) \frac{\sum\limits_{i=1}^{n} \frac{\lambda_i\eta_i^2}{(1-\theta\lambda_i)^3}}{\sum\limits_{i=1}^{n} \frac{(1+\lambda_i)\eta_i^2}{(1-\theta\lambda_i)^2}} - 2(1+\theta)(\sum\limits_{i=1}^{n} \frac{\lambda_i\eta_i^2}{(1-\theta\lambda_i)^3})(\frac{\sum\limits_{i=1}^{n} \frac{\eta_i^2}{1-\theta\lambda_i}}{\sum\limits_{i=1}^{n} \frac{(1+\lambda_i)\eta_i^2}{(1-\theta\lambda_i)^2}})^2$$

$$-4(1+\theta) \frac{\sum\limits_{i=1}^{n} \frac{\eta_i^2}{(1-\theta\lambda_i)^2}}{(\sum\limits_{i=1}^{n} \frac{(1+\lambda_i)\eta_i^2}{(1-\theta\lambda_i)^2})^2} (\sum\limits_{i=1}^{n} \frac{\lambda_i(1+\lambda_i)\eta_i^2}{(1-\theta\lambda_i)^3})$$

$$-2(1+\theta) \frac{\sum\limits_{i=1}^{n} \frac{\eta_i^2}{(1-\theta\lambda_i)^2}}{\{\sum\limits_{i=1}^{n} \frac{(1+\lambda_i)\eta_i^2}{(1-\theta\lambda_i)^2}\}^2} (\sum\limits_{i=1}^{n} \frac{\lambda_i\eta_i^2}{(1-\theta\lambda_i)^2}) (\sum\limits_{i=1}^{n} \frac{\eta_i^2}{(1-\theta\lambda_i)})$$

$$+4(1+\theta) \frac{(\sum\limits_{i=1}^{n} \frac{\eta_i^2}{(1-\theta\lambda_i)})^2}{(\sum\limits_{i=1}^{n} \frac{(1+\lambda_i)\eta_i^2}{(1-\theta\lambda_i)^2})^3} (\sum\limits_{i=1}^{n} \frac{\eta_i^2}{(1-\theta\lambda_i)^2}) (\sum\limits_{i=1}^{n} \frac{\lambda_i(1+\lambda_i)\eta_i^2}{(1-\theta\lambda_i)^3})$$

One can easily (plot $\Psi'(\theta)$ against θ and) check on which interval(s) of the real line the following condition is satisfied :

$$|\Psi'(\theta)| \leq k < 1, \tag{3.3.10}$$

or, equivalently see where $\Psi'(\theta)$ crosses the $\Psi'(\theta) = \pm 1$ lines. If on some interval $[a,b]$, (3.3.10) holds, then for any point θ_o of this interval the sequence of points $\theta_o, \theta_1, \ldots, \theta_i, \ldots$ where $\theta_{i+1} = \Psi(\theta_i)$, converges to the root of the equation $\theta = \Psi(\theta)$ in $[a,b]$ (see [33,65]).

Remark 3.3.4 : Noting the expression for $\Psi(\theta)$ in (3.3.9), it is clear that once we find a P which simultaneously diagonalizes R_1 and R_2, the inversion of matrices can be avoided to carry out the iteration under consideration.

An Illustrative Example

Example 3.3.2 Take $R_1 = \begin{bmatrix} 1 & .2 \\ .2 & 1 \end{bmatrix}$, $R_2 = \begin{bmatrix} 1 & -.5 \\ -.5 & 1 \end{bmatrix}$ and $\delta = \begin{bmatrix} 0.20 \\ 1.16 \end{bmatrix}$

We first find P which simultaneously diagonalizes R_1 and R_2 applying the transformation $\omega = P \chi$. The method of finding P is described in Remark 5.3.1, Chapter V.

We have,

$$C = \begin{bmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\[2ex] \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} \end{bmatrix}$$

$$D^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{1.2}} & 0 \\ 0 & \frac{1}{\sqrt{.8}} \end{bmatrix}$$

$$F = E'R_2E$$

$$= \begin{bmatrix} .41 & 0 \\ 0 & 1.87 \end{bmatrix}$$

Thus, $EH = \begin{bmatrix} .64 & .79 \\ .64 & -.79 \end{bmatrix}$ and consequently the

required $P = \begin{pmatrix} .64 & .79 \\ .64 & -.79 \end{pmatrix}^{-1} = \begin{pmatrix} .79 & .79 \\ .64 & -.64 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 0.41 & 0 \\ 0 & 1.87 \end{pmatrix}$

Now, $\underset{\sim}{\eta} = (P')^{-1}\underset{\sim}{\delta} = \begin{pmatrix} 0.8704 \\ -.7584 \end{pmatrix}$.

We note $A \neq 0$, A is defined in (3.3.8).

Hence we are in a position to compute $\Psi'(\theta)$.

We see that in $[-9,-8]$, $\Psi'(\theta)$ satisfies, $0 < \Psi'(\theta) < 1$. The graph of $\Psi'(\theta)$ is plotted in Fig. 3.1. Thus $\Psi(\theta)$ is a contraction mapping in $[-9,-8]$. We now verify this result by directly carrying out the iteration involved in (3.3.6). The following table shows this verification.
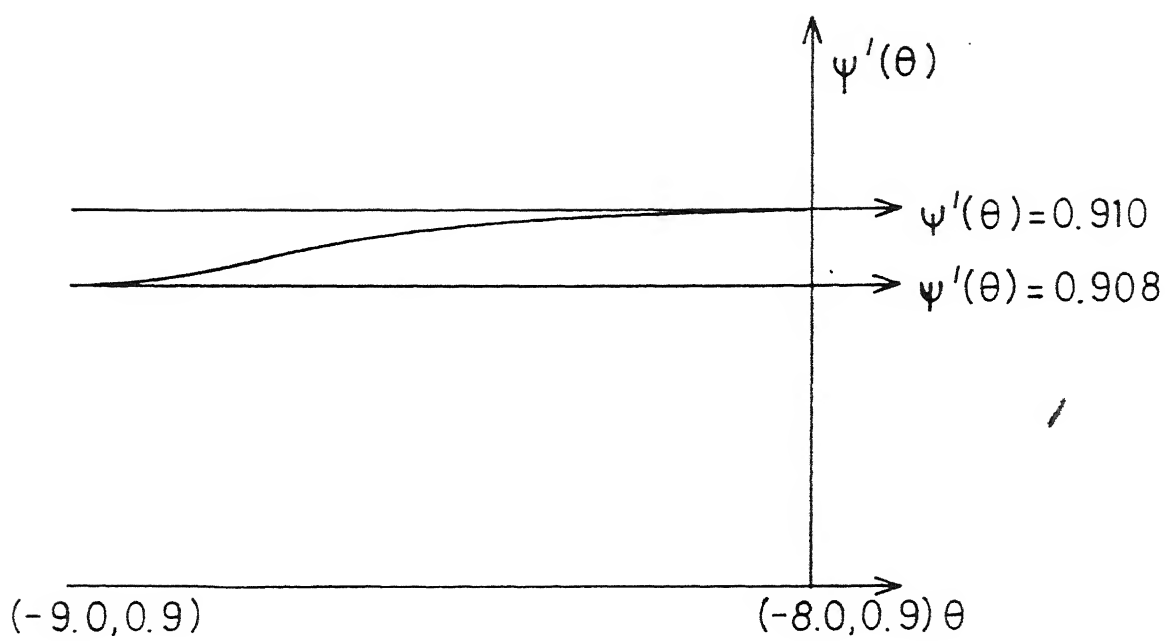
Graph of $\psi'(\theta)$

Fig. 3.1

<div align="center">

Table 3.1

Verification in Example 3.3.2

</div>

Initial Value θ_o	Final value θ_*	Number of iteration taken
-9.0	-8.9093	4
-8.5	-8.9093	5
-8.9	-8.9093	1

3.3.3 SOME SPECIAL CATEGORIES OF PROBLEMS

We first consider the following problem where the difference of the mean vectors of $\underset{\sim}{x}$ under H_1 and H_2 is the null vector

(1) As $\underset{\sim}{\delta} = 0$,

$$\rho_2(1,2;y) = \frac{\{(\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha})(\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha})\}^{\frac{1}{4}}}{\{\frac{1}{2}\underset{\sim}{\alpha}'(R_1+R_2)\underset{\sim}{\alpha}\}^{\frac{1}{2}}} \qquad (3.3.11)$$

Since there always exists a non-singular matrix P such that

$$R_1 = P'P$$

and $\qquad R_2 = P'\Lambda P$

where Λ is diagonal with elements as the eigen values of matrix $R_2 R_1^{-1}$, we can rewrite (3.3.11) as

$$\rho_2(1,2;y) = \frac{\sqrt{2}(\frac{\underset{\sim}{\beta}'\Lambda\underset{\sim}{\beta}}{\underset{\sim}{\beta}'\underset{\sim}{\beta}})^{1/4}}{(1 + \frac{\underset{\sim}{\beta}'\Lambda\underset{\sim}{\beta}}{\underset{\sim}{\beta}'\underset{\sim}{\beta}})^{1/2}} \qquad (3.3.12)$$

where $\qquad P\underset{\sim}{\alpha} = \underset{\sim}{\beta}.$ $\qquad$ (3.3.13)

Then we have the following theorem.

<u>Theorem 3.3.1</u> : The optimal $\underset{\sim}{\beta}$ is the eigen vector corresponding to $\lambda_{min}(R_2R_1^{-1})$ or $\lambda_{max}(R_2R_1^{-1})$ according as

$$\lambda_{min}(R_2R_1^{-1}) \ \lambda_{max}(R_2R_1^{-1}) \lessgtr 1 \qquad (3.3.14)$$

The optimal $\underset{\sim}{\alpha}$ is obtained by solving (3.3.13).

<u>Proof</u> : We see that (3.3.12) is of the form

$$y = \frac{x^{\frac{1}{4}}}{(1+x)^{\frac{1}{2}}}$$

where $x\varepsilon[a,b]$, and let $a < 1 < b$ (justification of which will be seen soon).

Then

$$\frac{dy}{dx} = \frac{1}{4} \frac{1-x}{(1+x)^{3/2}.x^{3/4}}$$

Thus y is increasing in $[a,1]$ and decreasing in $[1,b]$.

Now, we know that

$$\lambda_{min}(\Lambda) \leqq \frac{\underset{\sim}{\beta}'\Lambda\underset{\sim}{\beta}}{\underset{\sim}{\beta}'\underset{\sim}{\beta}} \leq \lambda_{max}(\Lambda) \ \forall \underset{\sim}{\beta}$$

where the equality at the left occurs when $\underset{\sim}{\beta}$ is the e.v. corresponding to the $\lambda_{min}(\Lambda)$ and at the right when $\underset{\sim}{\beta}$ is the e.v. corresponding to the $\lambda_{max}(\Lambda)$.

Thus,

$$\rho_2(1,2;y)\big|_{\hat{\beta}} = \min\left\{ \frac{\sqrt{2}(\lambda_{min}(\Lambda))^{\frac{1}{4}}}{(1+\lambda_{min}(\Lambda))^{\frac{1}{2}}} \,, \, \frac{\sqrt{2}(\lambda_{max}(\Lambda))^{\frac{1}{4}}}{(1+\lambda_{max}(\Lambda))^{\frac{1}{2}}} \right\}$$

$$(3.3.15)$$

Finally (3.3.14) follows easily if we observe that

$$\frac{(\lambda_{min}(\Lambda))^{\frac{1}{4}}}{(1+\lambda_{min}(\Lambda))^{\frac{1}{2}}} \lessgtr \frac{(\lambda_{max}(\Lambda))^{\frac{1}{4}}}{(1+\lambda_{max}(\Lambda))^{\frac{1}{2}}}$$

$$\Leftrightarrow \quad \frac{a_1}{(1+a_1)^2} \lessgtr \frac{b_1}{(1+b_1)^2} \,, \text{ where } a_1 = \lambda_{min}(\Lambda) < \lambda_{max}(\Lambda) = b_1$$

$$\Leftrightarrow a_1 + a_1 b_1^2 \lessgtr b_1 + b_1 a_1^2$$

$$\Leftrightarrow a_1 b_1 (b_1 - a_1) \lessgtr (b_1 - a_1)$$

$$\Leftrightarrow a_1 b_1 \lessgtr 1 \,.$$

Hence the theorem.

Next we consider the following problem.

(2) Let $\underset{\sim}{\delta} = 0$

and $R_1 = R_2 + \underset{\sim}{\mu}\underset{\sim}{\mu}'$ for some $\underset{\sim}{\mu} \neq 0$ (3.3.16)

This model occurs naturally in radar problems which we shall consider in Chapter VI . In this case the optimal $\underset{\sim}{\alpha}$ in the

sense of maximizing the Bhattacharyya distance is given by the following theorem.

__Theorem 3.3, 2__ : If the difference of the mean vectors of $\underset{\sim}{x}$ under H_1 and H_2 is null and R_1 and R_2 are related by (3.3.16), then the optimum $\underset{\sim}{\alpha}$ is given by

$$\underset{\sim}{\alpha} = R_2^{-1}\underset{\sim}{\mu} \; .$$

__Proof__ : It follows immediately once we recognize

$$\rho_2(1,2;y) = \frac{\sqrt{2}\{\underset{\sim}{\alpha}'(R_2+\underset{\sim}{\mu}\underset{\sim}{\mu}')\underset{\sim}{\alpha}\}^{\frac{1}{4}} (\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha})^{\frac{1}{4}}}{\{\underset{\sim}{\alpha}'(2R_2+\underset{\sim}{\mu}\underset{\sim}{\mu}')\underset{\sim}{\alpha}\}^{\frac{1}{2}}}$$

$$= \frac{\{1 + (\underset{\sim}{\alpha}'\underset{\sim}{\mu})^2/\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha}\}^{\frac{1}{4}}}{\{1+ \frac{1}{2} \frac{(\underset{\sim}{\alpha}'\underset{\sim}{\mu})^2}{\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha}}\}^{\frac{1}{2}}} \qquad (3.3.17)$$

as of the form $\quad y = \dfrac{(1+x)^{\frac{1}{4}}}{(1+ \frac{1}{2}x)^{\frac{1}{2}}}$

which decreases as x increases for $x > 0$ and thus our problem reduces to :

$$\max_{\underset{\sim}{\alpha}} (\underset{\sim}{\alpha}'\underset{\sim}{\mu})^2/(\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha}).$$

__Example 3.3.3.__ Take $R_2=\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$, $\underset{\sim}{\mu} = (\frac{1}{\sqrt{2}} , \frac{1}{\sqrt{2}})'$.

Thus, $R_1 = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$.

Since the optimal $\underset{\sim}{\alpha}$ is $R_2^{-1}\underset{\sim}{\mu}$,

$$\rho_2(1,2;y)\Big|_{\hat{\alpha}} = \frac{(1 + \underset{\sim}{\mu}' R_2^{-1}\underset{\sim}{\mu})^{\frac{1}{4}}}{(1 + \frac{1}{2}\underset{\sim}{\mu}' R_2^{-1}\underset{\sim}{\mu})^{\frac{1}{2}}} = 0.92$$

and from (3.3.15),

$$\rho_2(1,2;y)\Big|_{\hat{\beta}} = \min \left\{ \frac{\sqrt{2}(\lambda_{min}(\Lambda))^{\frac{1}{4}}}{(1+\lambda_{min}(\Lambda))^{\frac{1}{2}}} , \frac{\sqrt{2}(\lambda_{max}(\Lambda))^{\frac{1}{4}}}{(1 +\lambda_{max}(\Lambda))^{\frac{1}{2}}} \right\}$$

$$= \min \{0.92, 1.0\}$$

$$= .92.$$

Thus two results agree as expected.

3) Let $\underset{\sim}{\delta} = 0$

and $R_1 = R_2 + R_3$, where R_3 in a p.d. matrix . (3.3.18)

In Chapter VI, we shall describe a model which gives rise to the above relation between R_1 and R_2. In this case the following theorem states what the optimum $\underset{\sim}{\alpha}$ is.

Theorem 3.3.3: Let the difference of the mean vectors of $\underset{\sim}{x}$ under H_1 and H_2 be null and R_1 and R_2 be related according to (3.3.18). Then the required optimal $\underset{\sim}{\alpha}$ is the eigen vector corresponding to the maximum eigen value of $R_2^{-1}R_3$.

Proof : By virtue of (3.3.18),

$$\rho_2(1,2;y) = (1 + \frac{\alpha'R_3\alpha}{\alpha'R_2\alpha})^{\frac{1}{4}} / (1 + \frac{1}{2}\frac{\alpha'R_3\alpha}{\alpha'R_2\alpha})^{\frac{1}{2}}$$

(3.3.19)

Now, (3.3.19) is of the form

$$y = \frac{(1+x)^{\frac{1}{4}}}{(1+\frac{1}{2}x)^{\frac{1}{2}}}$$

which decreases as x increases for x > 0.

Thus our problem reduces to find

$$\max_{\alpha} \frac{\alpha'R_3\alpha}{\alpha'R_2\alpha}$$

(3.3.20)

Hence the theorem.

3.3.4 OTHER LINEAR DISCRIMINANT FUNCTIONS

Various other distances have been considered by Kullback in the context of linear discriminant functions ([31]). The following expressions are of interest :

$$I(1,2;y) = \frac{1}{2} \ln \frac{\alpha'R_2\alpha}{\alpha'R_1\alpha} - \frac{1}{2} + \frac{1}{2} \frac{\alpha'R_1\alpha}{\alpha'R_2\alpha} + \frac{1}{2} \frac{(\alpha'\delta)^2}{\alpha'R_2\alpha}$$

(3.3.21)

$$I(2,1;y) = \frac{1}{2}\ln\frac{\alpha'R_1\alpha}{\alpha'R_2\alpha} - \frac{1}{2} + \frac{1}{2}\frac{\alpha'R_2\alpha}{\alpha'R_1\alpha} + \frac{1}{2}\frac{(\alpha'\delta)^2}{\alpha'R_1\alpha}$$

$$(3.3.22)$$

$$J(1,2;y) = \frac{1}{2}\frac{\alpha'R_2\alpha}{\alpha'R_1\alpha} + \frac{1}{2}\frac{\alpha'R_1\alpha}{\alpha'R_2\alpha} - 1 + \frac{1}{2}\left(\frac{1}{\alpha'R_1\alpha} + \frac{1}{\alpha'R_2\alpha}\right)(\alpha'\delta)^2$$

$$(3.3.23)$$

The value of α for which $I(1,2;y)$ in (3.3.21) is a maximum satisfies (by the usual calculus procedures) an equation of the same form as (3.3.6) but with

$$-\theta = \frac{\alpha'R_1\alpha}{\alpha'R_2\alpha}\left(1 - \frac{(\alpha'\delta)^2}{\alpha'R_2\alpha-\alpha'R_1\alpha}\right) \qquad (3.3.24)$$

The value of α for which $I(2,1;y)$ in (3.3.22) is a maximum satisfies (by the usual calculus procedures) an equation of the same form as (3.3.6) but with

$$-\theta = \frac{\alpha'R_1\alpha(\alpha'R_1\alpha-\alpha'R_2\alpha)}{\alpha'R_2\alpha(\alpha'R_1\alpha-\alpha'R_2\alpha-(\alpha'\delta)^2)} \qquad (3.3.25)$$

The value of α for which $J(1,2;y)$ in (3.3.23) is a maximum satisfies (by the usual calculus procedures) an equation of the same form as (3.3.6) but with

$$-\theta = \frac{\alpha'R_1\alpha((\alpha'R_2\alpha)^2-(\alpha'R_1\alpha)^2-(\alpha'\delta)^2(\alpha'R_1\alpha))}{\alpha'R_2\alpha((\alpha'R_2\alpha)^2-(\alpha'R_1\alpha)^2+(\alpha'\delta)^2(\alpha'R_2\alpha))}$$

$$(3.3.26)$$

Another important class of linear discriminant functions arises in the following way. It is clear that associated with the classification scheme

$$\alpha'x \underset{H_2}{\overset{H_1}{\gtrless}} c$$

$$(\text{accept})$$

are two kinds of errors. The probability of misclassifying an observation when it comes from the population under H_1 is given by

$$e_1 = Pr_1(\alpha'x < c) = Pr_1\left(\frac{\alpha'x - \alpha'\mu_1}{(\alpha'R_1\alpha)^{1/2}} < \frac{c - \alpha'\mu_1}{(\alpha'R_1\alpha)^{1/2}}\right)$$

$$= \phi\left(\frac{c - \alpha'\mu_1}{(\alpha'R_1\alpha)^{1/2}}\right) = 1 - \phi\left(\frac{\alpha'\mu_1 - c}{(\alpha'R_1\alpha)^{1/2}}\right)$$

$$(3.3.27)$$

and the probability of misclassifying an observation when it comes from the population under H_2 is given by

$$e_2 = Pr_2(\alpha'x > c) = 1 - Pr_2(\alpha'x < c)$$

$$= 1 - \phi\left(\frac{c - \alpha'\mu_2}{(\alpha'R_2\alpha)^{1/2}}\right)$$

$$(3.3.28)$$

Define $y_1 = \dfrac{\alpha'\mu_1 - c}{(\alpha'R_1\alpha)^{1/2}}$ and $y_2 = \dfrac{c - \alpha'\mu_2}{(\alpha'R_2\alpha)^{1/2}}$ $\qquad(3.3.29)$

We can then form a minimum error criterion for finding a linear discriminant function, namely, for a given e_1, what

linear function of the x's will minimize e_2 ? Since e_1 and e_2 are monotone functions of y_1 and y_2 respectively, it is simpler to work with the latter.

y_2 can be rewritten as

$$y_2 = \frac{\alpha'\delta - y_1(\alpha'R_1\alpha)^{1/2}}{(\alpha'R_2\alpha)^{1/2}} \tag{3.3.30}$$

Thus for a given e_1, e_2 will be minimized by maximizing (3.3.30). The usual calculus procedures lead to the equation

$$\alpha = (R_1 - \theta R_2)^{-1}\delta$$

where

$$-\theta = (\frac{y_2}{y_1})(\frac{\alpha'R_1\alpha}{\alpha'R_2\alpha})^{1/2} , \tag{3.3.31}$$

(see Kullback([31])).

<u>Remark 3.3.5</u> : We note that the linear discriminant functions derived from the minimum error criterion were extensively studied by Anderson and Bahadur ([4]).

<u>Remark 3.3.6</u> : If θ found by maximizing $-\ln p_2(1,2;y)$ makes R_θ positive definite and the threshold of the test (3.2.2) c is chosen according to the following relation

$$c = \delta'R_\theta^{-1}\mu_1 - \delta'R_\theta^{-1}R_1R_\theta^{-1}\delta$$

then <u>our procedure</u> is <u>admissible</u> (within the class of linear procedures). This follows from the Anderson-Bahadur's theorem on admissible class of linear procedures (see <u>Remark 3.6.3</u>

where we describe it in some detail). All other procedures
considered above also become admissible following the same
approach and that the A–B procedure is admissible is proven
in ([4]).

3.3.5   COMPARISON OF THE VARIOUS LINEAR DISCRIMINANT FUNCTIONS

First we consider the example of Section 3.3.1.
Having done this we discuss examples based on autoregressive
processes.

(a) The LDFs listed below are obtained in([31])
for the example given in **Section 3.3.1** :

(number of iteration taken is 3)

$$\max I(1,2;y) : \quad y = x_1 - 0.3924x_2 \qquad (3.3.32)$$

$$\max I(2,1;y) : \quad y = x_1 - 0.8491x_2 \qquad (3.3.33)$$

$$\max J(1,2;y) : \quad y = x_1 - 0.6295x_2 \qquad (3.3.34)$$

$\max y_2$ (given $y_1 = 1.645$ or equivalently $e_1 = 0.05$):

$$y = x_1 - 0.4173x_2 \qquad (3.3.35)$$

$\max y_2$(given $y_2 = 1.0$ or equivalently $e_1 = 0.16$):

$$y = x_1 - 0.3990x_2 \qquad (3.3.36)$$

From **Example 3.3.1** , we have obtained,

$$\max(-\ln \rho_2(1,2;y)): y = x_1 - 0.4153x_2 \qquad (3.3.37)$$

In <u>Table 3.2</u> the errors of misclassification of
one kind given the other kind, that result due to the
use of the above linear discriminant functions are presented.

The results of all columns excepting the first are collected from Kullback ([31]). The linear discriminant function in the last column is found by pooling variances and covariances between the samples and proceeding as if the covariance matrices were the same.

It is clear from Table 3.2 that maximizing $I(2,1;y)$ and $J(1,2;y)$ yields the LDFs which have larger errors of misclassification than the other five for whom the errors of misclassification are very much alike.

(b) In this section we present some comparisons for a class of problems where the basic process under consideration follows an autoregressive scheme. In the following examples, we study the convergence and the rate or convergence of the iteration process involved (3.3.6, 3.3.24, 3.3.25, 3.3.26, 3.3.31) and also compute the type II error e_2 resulting from the use of the LDFs obtained by maximizing the distances and by the Anderson-Bahadur (denoted by A-B) procedure for a given type I error e_1. We consider the underlying process as to be a non-stationary autoregressive (AR) process of order one and two. We divide the examples into two parts depending upon the order of the AR process.

(i) A first order AR scheme is defined as

$$Z(t+1) = \rho Z(t) + \varepsilon(t+1), \quad (t = 0,1,..,n-1,...) \qquad (3.3.38)$$

where $\{ Z(t), t \geq 0 \}$ is the centered process i.e.
$Z(t) \triangleq X(t) - \mu(t)$

and

$$E\varepsilon(t) = 0 \qquad \forall\ t$$

$$E\varepsilon(t)\varepsilon(t+\tau) = \begin{cases} 0 & \text{if } \tau \neq 0 \\ 1 & \text{if } \tau = 0 \end{cases}$$

By repeated substitution of

$$Z(t-i) = \rho\ Z(t-i-1) + \varepsilon(t-i) \text{ for } i = 0,..,n-1$$

in (3.3.38), we obtain,

$$Z(t+1) = \rho^n\ Z(t+1-n) + \sum_{i=o}^{n-1} \rho^i\ \varepsilon\ (t+1-i)$$

or $\qquad Z(t) = \rho^n\ Z(t-n) + \sum_{i=o}^{n-1} \rho^i\ \varepsilon\ (t-i)$ $\hfill$ (3.3.38)'

Assume $Z(o) \equiv 0$.

Then (3.3.38)' reduces to (see Bhat [10])

$$Z(n) = \sum_{i=o}^{n-1} \rho^i\ \varepsilon(n-i)$$

Thus $EZ(n)Z(n+\tau) = E(\sum_{i=o}^{n-1} \rho^i\ \varepsilon(n-i)(\sum_{i=o}^{n+\tau-1} \rho^i\ \varepsilon\ (n+\tau-i))$

$$= \sum_{i=o}^{n-1} \sum_{j=o}^{n+\tau-1} \rho^i\ \rho^j E\ \varepsilon\ (n-i)\varepsilon(n+\tau-j)$$

$$= \sum_{i=o}^{n-1} \rho^i\ \rho^{i+\tau}$$

$$= \rho^\tau \sum_{i=o}^{n-1} (\rho^2)^i$$

$$= \frac{1-\rho^{2n}}{1-\rho^2}\ \rho^\tau$$

i.e. $EZ(i)Z(j) = \dfrac{1-\rho^{2i}}{1-\rho^2}\ \rho^{(j-i)}$ $\hfill$ (3.3.39)

We determine $\{\mu(t), t \geq 0\}$ to satisfy the first order difference equation :

$$\mu(t+1) = \rho\,\mu(t)$$

(3.3.40)

Now we consider the following examples.

Example 3.3.4 : Under H_1 : $\rho = .9$, $\mu_0 \equiv 1$

Under H_2 : $\rho = -.9$, $\mu_0 \equiv 0$

Example 3.3.5 : Under H_1 : $\rho = .2$

Under H_2 : $\rho = .7$

Example 3.3. 6 : Under H_1 : $\rho = .2$

Under H_2 : $\rho = .15$

The δ vector is the same in all the above examples where the sample size varies from 2 to 20. The distinguishing feature of the examples is that the ρ's differ significantly in Example 3.3.4, differ moderately in Example 3.3.5., and are very close in the third Example 3.3.6 . The computations are shown in the following tables. The initial value of θ in the iterations is -1 in all the examples, n denotes the number of observations and " iter " denotes the number of iterations required to get the optimal θ (denoted by $\hat{\theta}$). The comparisons of the performances of the LDFs are also shown graphically in the figures. We first present the graphs to have an over-all view.
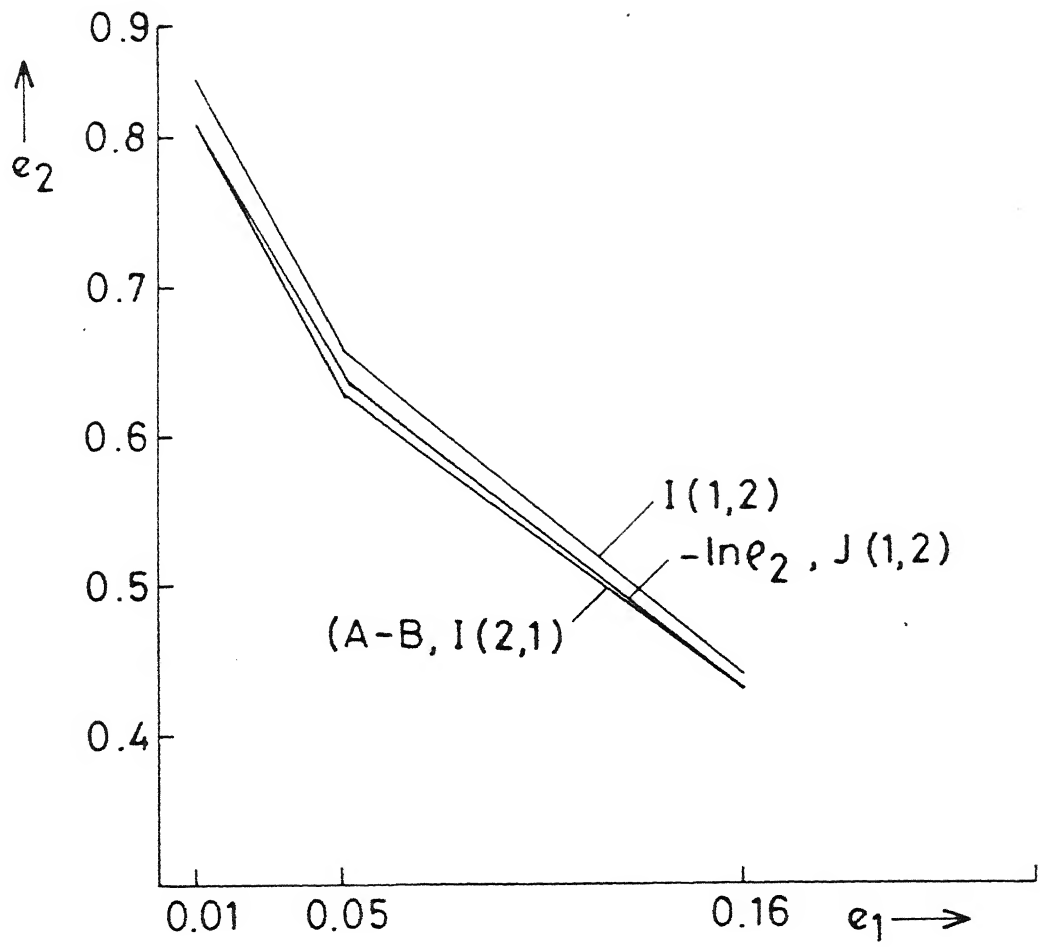
Fig. 3.2 Example 3.3.5 (AR (1)), n = 2
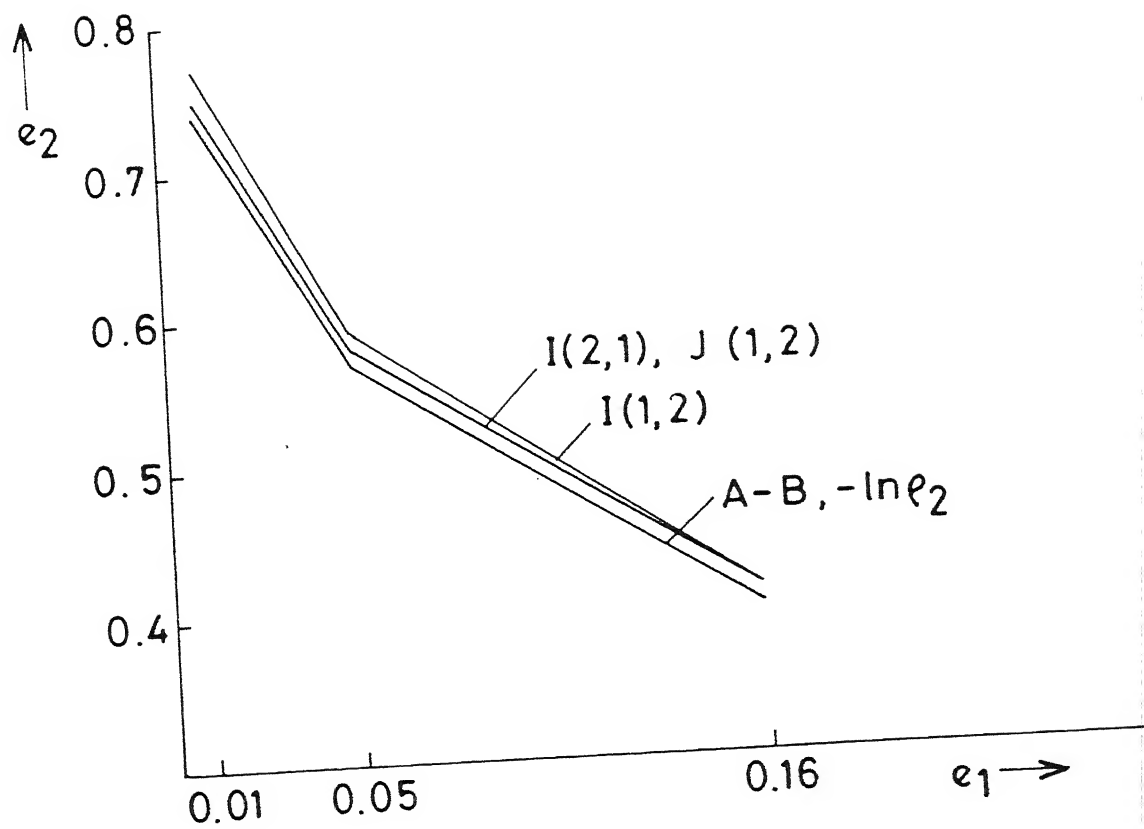
Fig.3.3 Example 3.3.5(AR(1)), n = 3
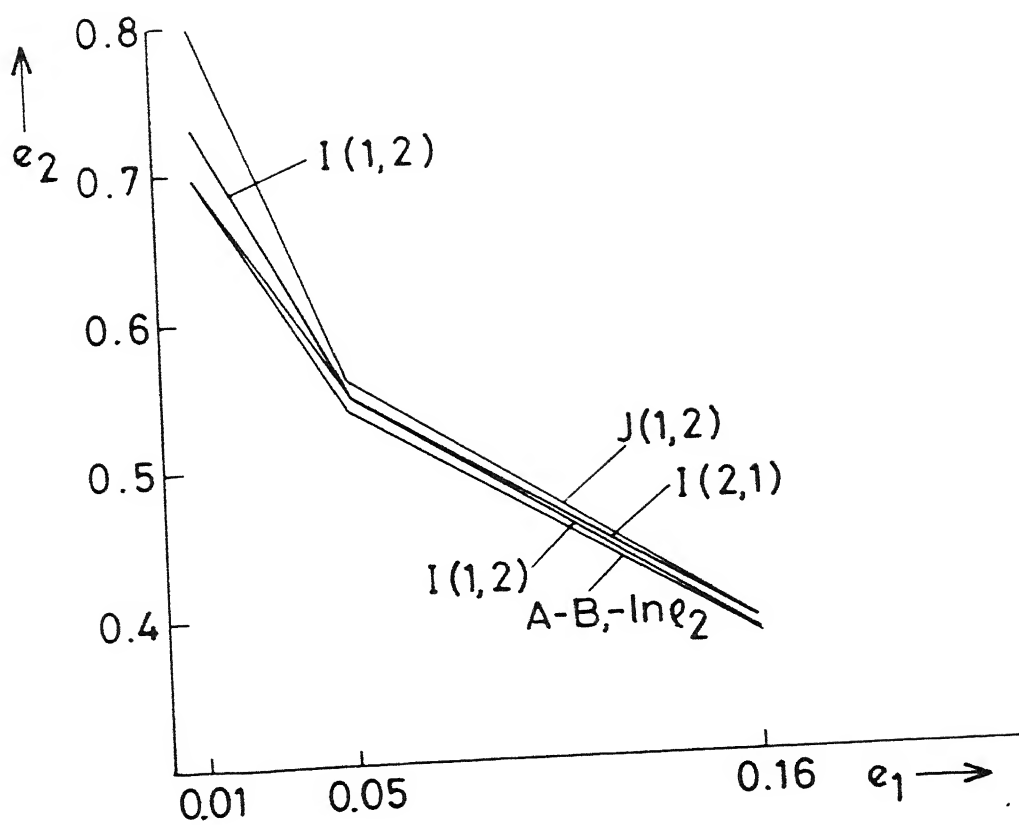
Fig.3.4 Example 3.3.5 (AR(1)), n=4
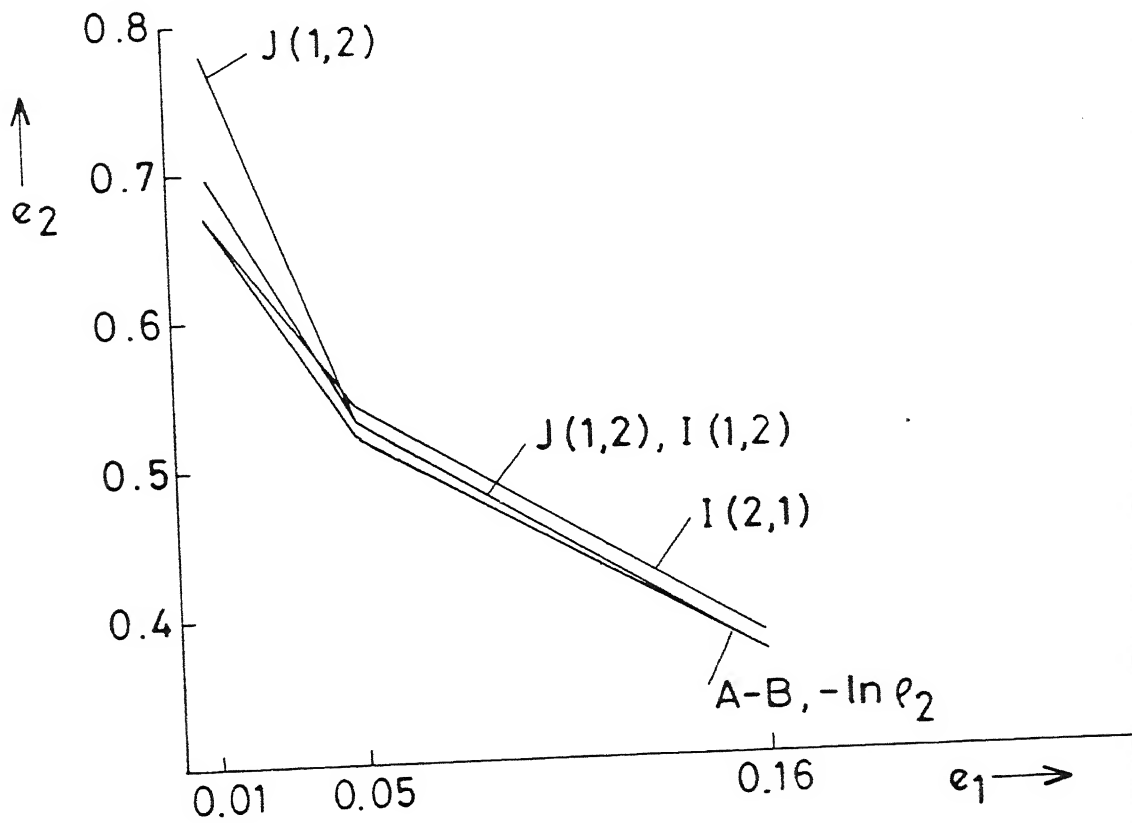
Fig.3.5 Example 3.3.5 (AR (1)). n= 5
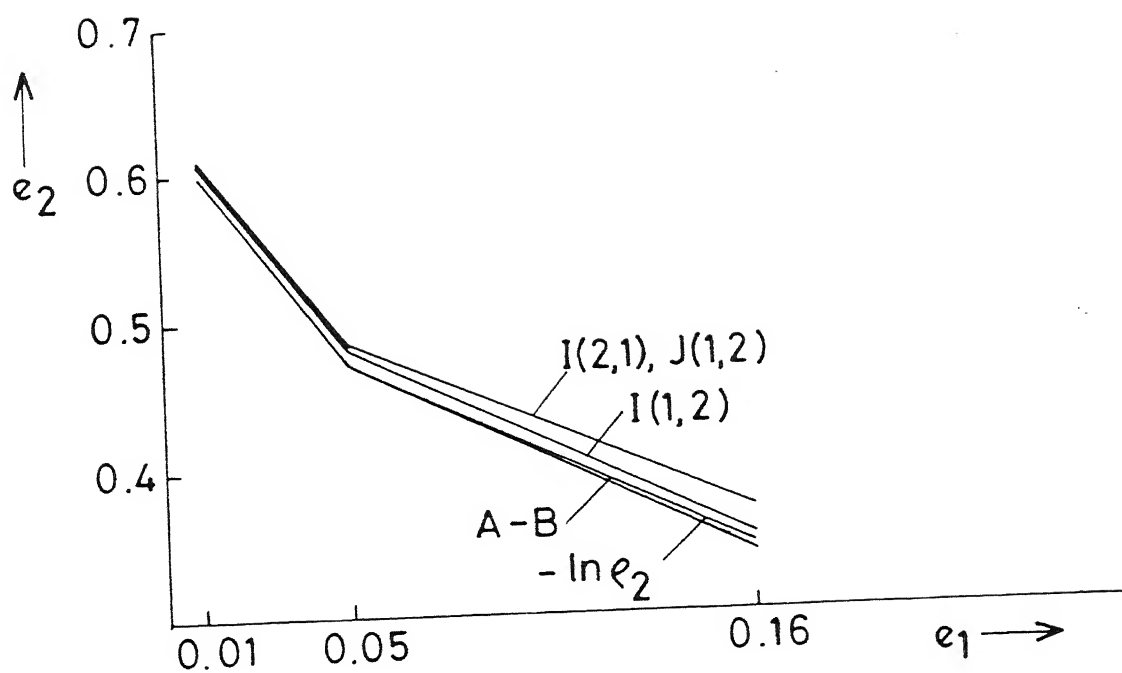
Fig.3.6 Example 3.3.5 (AR (1)), n =10
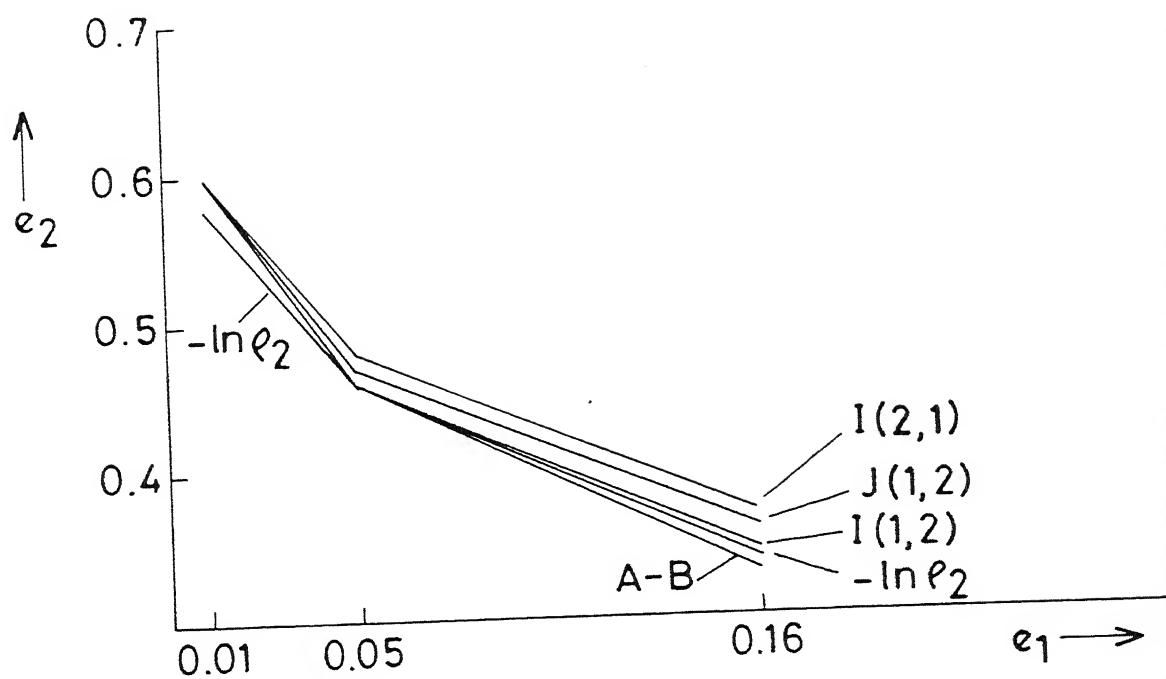
Fig.3.7 Example 3.3.5 (AR (1)). n = 20
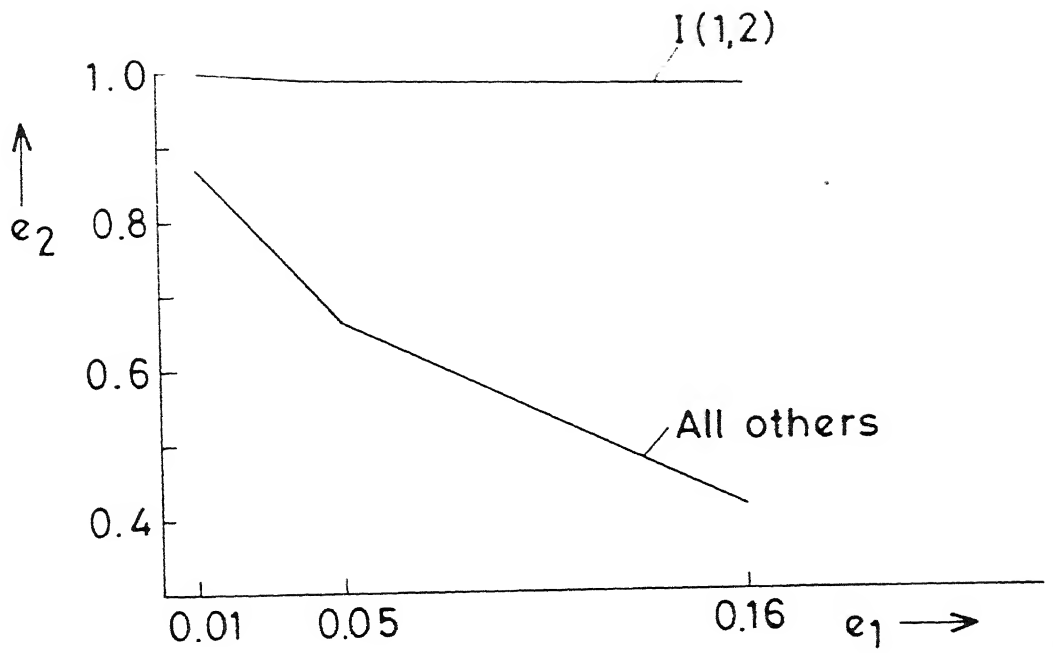
Fig. 3.8 Example 3.3.6(AR(1)), n=2

Table 3.3

Type II errors e_2 resulting from the use of LDFs obtained
by maximizing $-\ln\rho_2(1,2;y)$

Example	$\hat{\theta}$	iter	n	$e_1 = .01$	$e_1 = .05$	$e_1 = .16$
3.3.4	49.5765	5	2	1.0	1.0	1.0
	31.6438	6	3	1.0	1.0	1.0
	36.2683	6	4	1.0	1.0	1.0
	43.6417	6	5	1.0	1.0	1.0
	89.8485	6	10	1.0	1.0	1.0
	57.5591	8	20	1.0	1.0	1.0
3.3.5	−.2252	4	2	0.8159	0.6368	0.4286
	−.0707	5	3	0.7486	0.5753	0.3974
	−.0153	5	4	0.7019	0.5398	0.3783
	−.0101	5	5	0.6700	0.5160	0.3669
	.0396	3	10	0.6000	0.4700	0.3450
	.0457	3	20	0.5800	0.4600	0.3400
3.3.6	−1.1327	2	2	0.8708	0.6664	0.4090
	−1.1480	2	3	0.8365	0.6104	0.3446
	−1.1455	2	4	0.8078	0.5636	0.3050
	−1.1403	2	5	0.7852	0.5279	0.2743
	−1.1223	2	10	0.7231	0.4701	0.2265
	−1.1148	2	20	0.6800	0.4000	0.1800

Table 3.4

Type II errors e_2 resulting from the use of LDFs obtained by maximizing $I(1,2;y)$

Example	$\hat{\theta}$	iter	n	$e_1=.01$	$e_1=.05$	$e_1=.16$
	10.3138	3	2	1.0	1.0	1.0
	18.3408	4	3	1.0	1.0	1.0
3.3.4	27.3149	5	4	1.0	1.0	1.0
	36.8808	6	5	1.0	1.0	1.0
	18.1935	6	10	1.0	1.0	1.0
	12.8295	7	20	1.0	1.0	1.0
	−2.0212	7	2	0.8365	0.6554	0.4404
	−0.7370	7	3	0.7734	0.5910	0.4090
3.3.5	−0.3578	9	4	0.7257	0.5517	0.3783
	−0.2012	9	5	0.6879	0.5239	0.3669
	−0.0305	9	10	0.6100	0.4800	0.3500
	−0.0020	10	20	0.6000	0.4600	0.3450
	34.5554	3	2	0.9998	0.9983	0.9884
	31.7064	3	3	0.9998	0.9990	0.9929
3.3.6	32.7184	3	4	1.0000	0.9993	0.9951
	33.9604	3	5	1.0000	0.9997	0.9967
	38.8674	3	10	1.0000	1.0000	1.0000
	41.3886	3	20	1.0000	1.0000	1.0000

Table 3.5

Type II errors e_2 resulting from the use of LDFs obtained by maximizing $I(2,1;y)$

Example	$\hat{\theta}$	iter	n	$e_1=.01$	$e_1=.05$	$e_1=.16$
	−83.2602	6	2	1.0	1.0	1.0
	36.0286	6	3	1.0	1.0	1.0
	37.3797	6	4	1.0	1.0	1.0
3.3.4	44.1045	6	5	1.0	1.0	1.0
	89.8911	8	10	1.0	1.0	1.0
	57.0704	6	20	1.0	1.0	1.0
	0.1781	3	2	0.8106	0.6331	0.4325
	0.1776	3	3	0.7454	0.5793	0.4052
	0.1630	3	4	0.6985	0.5438	0.3897
3.3.5	0.1499	3	5	0.6700	0.5398	0.3821
	0.1176	4	10	0.6100	0.4850	0.3700
	0.1056	4	20	0.6000	0.4800	0.3700
	−0.0335	2	2	0.8708	0.6664	0.4090
	−0.0376	2	3	0.8365	0.6104	0.3446
	−0.0369	2	4	0.8078	0.5636	0.3050
3.3.6	−0.0359	2	5	0.7852	0.5279	0.2743
	−0.0322	2	10	0.7231	0.4701	0.2265
	−0.0304	2	20	0.6800	0.4000	0.1800

Table 3.6

Type II errors e_2 resulting from the use of LDFs obtained by maximizing $J(1,2;y)$

Example	$\hat{\theta}$	iter	n	$e_1=.01$	$e_1=.05$	$e_1=.16$
	11.4294	4	2	1.0	1.0	1.0
	18.7921	4	3	1.0	1.0	1.0
3.3.4	27.5608	5	4	1.0	1.0	1.0
	37.0316	6	5	1.0	1.0	1.0
	18.2201	6	10	1.0	1.0	1.0
	12.8339	4	20	1.0	1.0	1.0
	−0.0156	4	2	0.8159	0.6368	0.4286
	0.1019	4	3	0.7486	0.5753	0.3974
3.3.5	0.1228	4	4	0.7454	0.5398	0.3859
	0.1243	4	5	0.6985	0.5160	0.3783
	0.1089	4	10	0.6100	0.4850	0.3700
	0.1004	3	20	0.6000	0.4700	0.3600
	−1.2462	2	2	0.8708	0.6664	0.4090
	−1.3145	2	3	0.8365	0.6104	0.3446
3.3.6	−1.3367	2	4	0.8078	0.5636	0.3080
	−1.3492	2	5	0.7852	0.5279	0.2743
	−1.3636	2	10	0.7231	0.4701	0.2265
	−1.3649	2	20	0.6800	0.4000	0.1800

Table 3.7

Type II errors e_2 resulting from the use of LDFs obtained by the A—B procedure

Example	$\hat{\theta}$	iter	n	$e_1=.01$	$e_2=.05$	$e_1=.16$
3.3.4	2.2411	4	2	1.0	1.0	0.6484
	5.8541	11	3	1.0	1.0	0.5000
	1.2776	4	4	1.0	1.0	0.5000
	2.2267	9	5	1.0	1.0	0.5000
	0.0000	10	10	1.0	1.0	0.5000
	−0.0020	8	20	1.0	1.0	0.5000
3.3.5	0.1639	3	2	0.8106	0.6331	0.4286
	0.0786	3	3	0.7454	0.5753	0.3974
	0.0378	3	4	0.6985	0.5359	0.3783
	0.0152	3	5	0.6700	0.5160	0.3669
	0.0546	4	10	0.6000	0.4700	0.3400
	−0.2392	4	20	0.5800	0.4600	0.3300
3.3.6	0.2704	2	2	0.8708	0.6664	0.4090
	0.1761	2	3	0.8365	0.6104	0.3446
	0.1039	2	4	0.8078	0.5636	0.3080
	0.0483	2	5	0.7852	0.5279	0.2743
	−0.8701	2	10	0.7231	0.4701	0.2265
	−0.9715	2	20	0.6800	0.4000	0.1800

We can draw the following conclusions from the above Tables.

(1) In **Example** 3.3.4, all the LDFs resulting from maximizing $-\ln\rho_2(1,2;y)$, $I(1,2;y)$, $I(2,1;y)$, $J(1,2;y)$ and that due to the Anderson-Bahadur (A-B) procedure accept H_1 except the LDF obtained by the A-B procedure when the type I error is .16 is which case Type II error is nearly .5000. The increase in the sample size up to 20 has no effect on the performances.

(2) In **Example** 3.3.5, the LDFs obtained by maximizing $-\ln\rho_2(1,2;y)$ do better than the others from the error point of view and its performance is the same as that of the A-B procedure in almost all cases. As n increases, the performances improve.

(3). In **Example** 3.3.6, all the LDFs have the same performance except the LDF due to $I(1,2;y)$ which does not do well. Here also, the performances improve as n increases.

(4) (i) In **Example** 3.3.4, the convergence of the iteration in maximizing $J(1,2;y)$, $I(1,2;y)$ is most rapid.

(ii) In **Example** 3.3.5, the number of iterations required to get an optimal θ is highest in maximizing $I(1,2;y)$ and the lowest in the A-B procedure and maximizing $I(2,1;y)$.

(iii) The number of iteration taken in maximizing $-\ln\rho_2(1,2;y)$, $I(2,1;y)$, $J(1,2;y)$ and in the A-B procedure is the same in **Example** 3.3.6 , which is lower than that of maximizing $J(1,2;y)$.

(ii) A second order AR scheme is given by

$$Z(t+2) = \beta_1 \, Z(t+1) + \beta_2 \, Z(t) + \varepsilon(t+2) \quad (t = 0,1,\ldots,n-1,\ldots)$$

$$(3.3.41)$$

Define an operator F by

$$F \, Z(t) \overset{\Delta}{=} Z(t+1)$$

Thus $(3.3.41)$ can be written as

$$(F^2 - \beta_1 F - \beta_2) \, Z(t) = \varepsilon(t+2)$$

The homogeneous part of this equation is

$$(F^2 - \beta_1 F - \beta_2) \, Z(t) = 0,$$

which has a general solution

$$Z(t) = \lambda \xi_1^t + \mu \xi_2^t$$

where ξ_1, ξ_2 are the solutions of the operator equation known as the characteristic equation :

$$F^2 - \beta_1 F - \beta_2 = 0.$$

We have,

$$\xi_1 = \frac{\beta_1 + (\beta_1^2 + 4\beta_2)^{1/2}}{2}$$

$$\xi_2 = \frac{\beta_1 - (\beta_1^2 + 4\beta_2)^{1/2}}{2}$$

Assume $\beta_1^2 + 4\beta_2 > 0$, so that ξ_1 and ξ_2 real numbers. A particular solution is obtained by writing

$$Z(t) = [(F-\xi_1)^{-1}(F-\xi_2)^{-1}] \varepsilon \, (t+2)$$

$$= [(1-\frac{\xi_1}{F})^{-1} (1 - \frac{\xi_2}{F})^{-1}] \varepsilon \, (t)$$

$$= \left[\sum_0^\infty (\frac{\xi_1}{F})^r \sum_0^\infty (\frac{\xi_2}{F})^s \right] \varepsilon(t)$$

$$= \sum_{r=0}^\infty \sum_{s=0}^\infty \xi_1^r \, \xi_2^s \, \varepsilon(t-r-s).$$

The complete solution of the difference equation (3.3.41) is the sum of the general solution of the homogeneous part and a particular solution.

Thus, we get

$$Z(t) = \lambda \xi_1^t + \mu \, \xi_2^t + \sum_{r=0}^\infty \sum_{s=0}^\infty \xi_1^r \xi_2^s \, \varepsilon(t-r-s)$$

Assume $Z(o) = Z(1) = 0$.

Thus λ and μ can be obtained as

$$\lambda = - \sum_{r=0}^\infty \xi_1^r \, \varepsilon_{1-r} / (\xi_1 - \xi_2)$$

$$\mu = \sum_{s=0}^\infty \xi_2^s \, \varepsilon_{1-s} / (\xi_1 - \xi_2)$$

Hence, after doing considerable simplification, we obtain (see Bhat [10])

$$Z(t) = \sum_{i=0}^{t-1} a_i \varepsilon(t-i)$$

where $a_i = \dfrac{\xi_1^{i+1} - \xi_2^{i+1}}{\xi_1 - \xi_2}$

Finally,

$$EZ(t)Z(t+h) = E\left(\sum_{i=0}^{t-1} a_i \varepsilon(t-i)\right)\left(\sum_{i=0}^{t+h-1} a_i \varepsilon(t+h-i)\right)$$

$$= \sum_j \sum_i a_i a_j \, E\varepsilon(t-i)\varepsilon(t+h-j)$$

$$= \sum_{i=0}^{t-1} a_i a_{i+h}$$

$$= \frac{1}{(\xi_1 - \xi_2)^2}\left[\frac{(1-\xi_1^{2t})\xi_1^{h+2}}{1-\xi_1^2} - \xi_1\xi_2(\xi_1^h + \xi_2^h)\frac{(1-\xi_1^t\xi_2^t)}{1-\xi_1\xi_2}\right.$$

$$\left. + \frac{(1-\xi_2^{2t})\xi_2^{h+2}}{1-\xi_2^2}\right] \qquad (3.3.42)$$

We determine $\{\mu(t)\}$ as a solution of the difference equation :

$$\mu(t+2) = \beta_1\mu(t+1) + \beta_2\mu(t).$$

Example 3.3.7 : Under H_1 : $\beta_1 = .7$, $\beta_2 = .1$, $\mu_o = 1$, $\mu_1 = 0$

Under H_2 : $\beta_1 = -.7$, $\beta_2 = -.1$, $\mu_o = \mu_1 = 0$

Example 3.3.8 : Under H_1 : $\beta_1 = .7$, $\beta_2 = .1$,

Under H_2 : $\beta_1 = .1$, $\beta_2 = .01$

Example 3.3.9 : Under H_1 : $\beta_1 = .7$, $\beta_2 = .1$

Under H_2 : $\beta_1 = .65$, $\beta_2 = .07$

The $\underset{\sim}{\delta}$ vector is the same in all the examples above.

We furnish the computations similar to those as in the AR(1) process-examples in the following Tables.
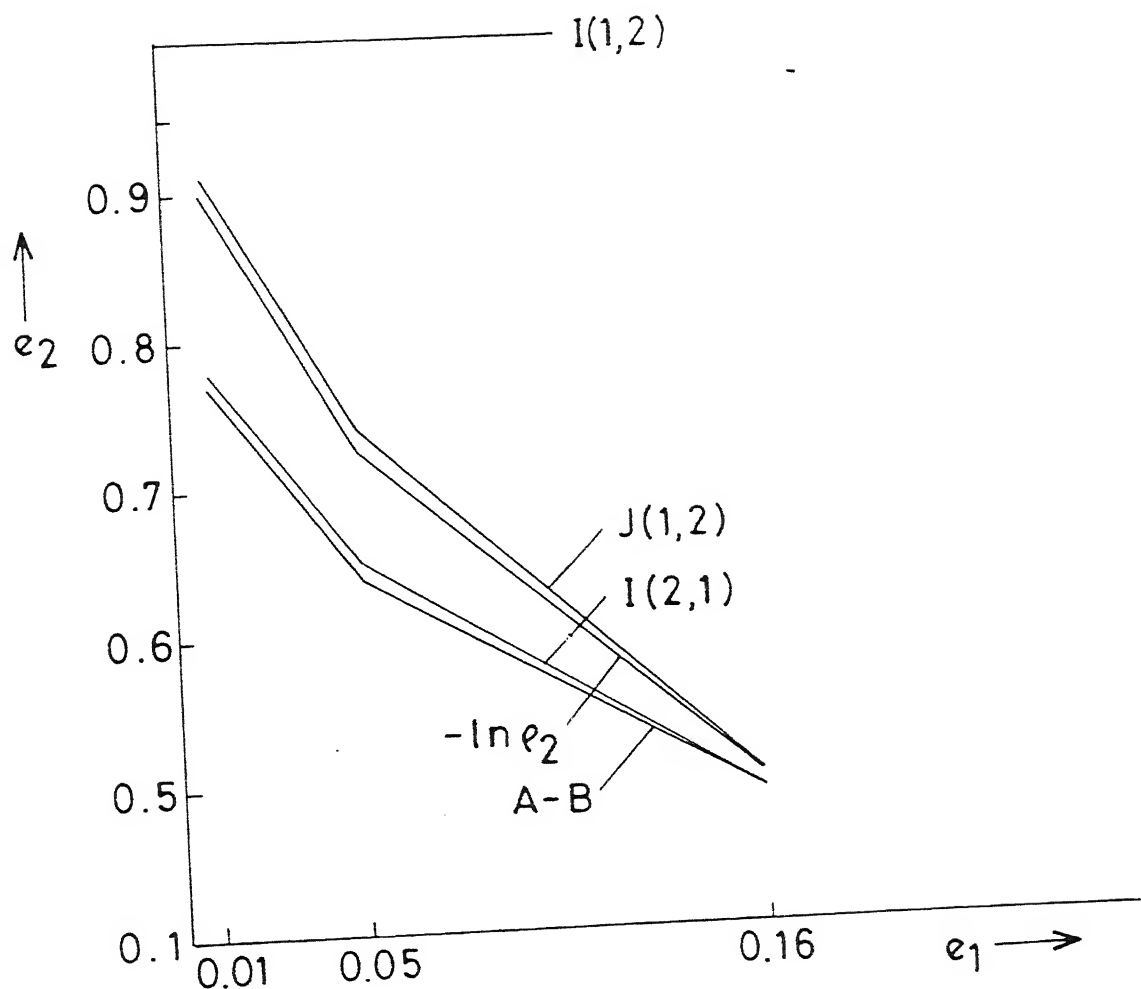
Fig. 3.10 Example 3.3.7 (AR (2)), n = 2
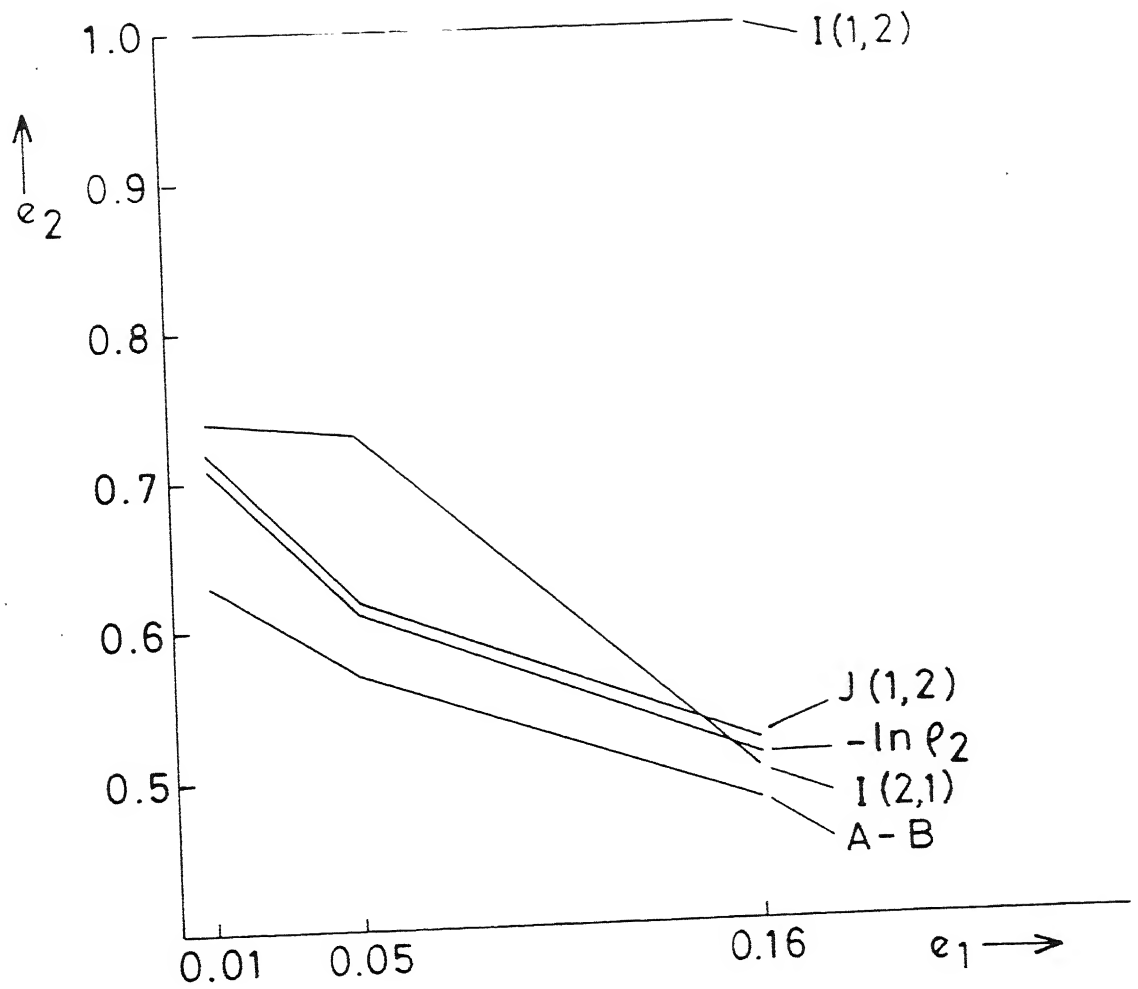
Fig.3.11 Example 3.3.7(AR(2)) n = 3
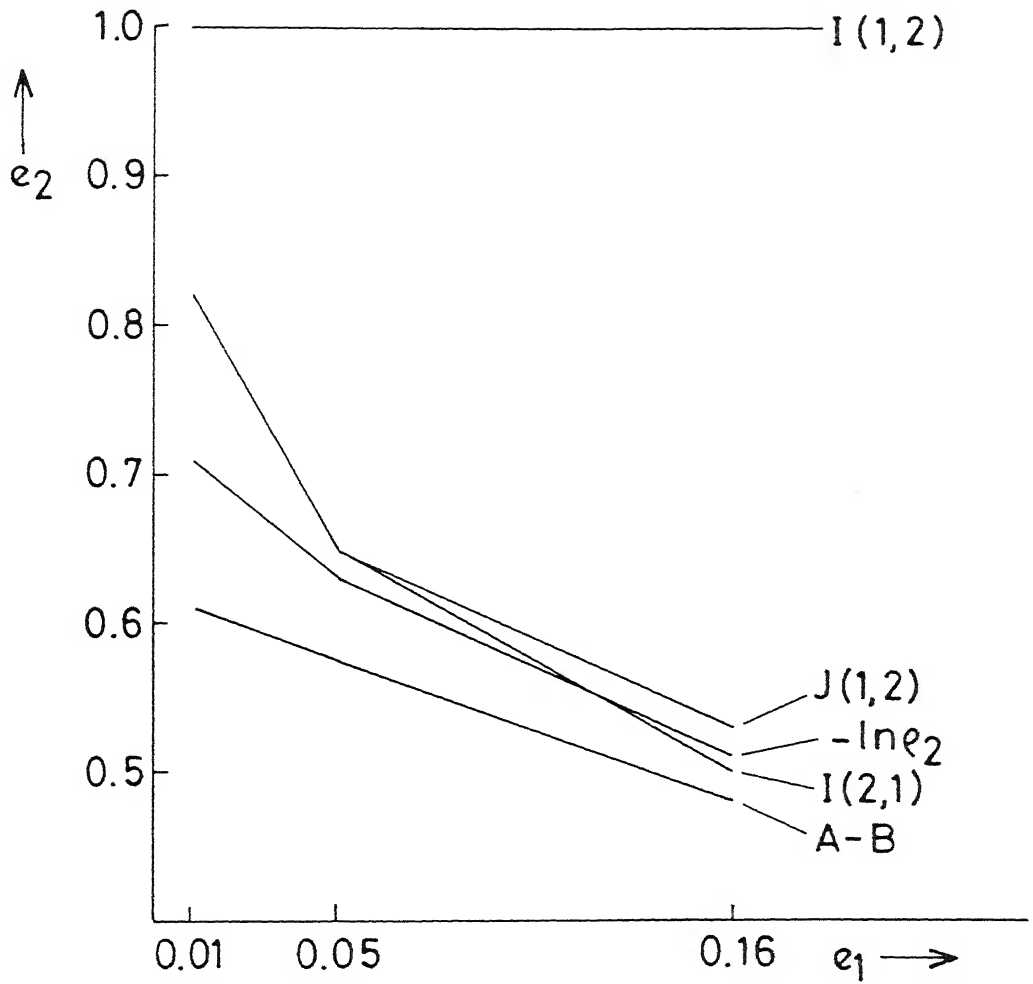
Fig.3.12 Example 3.3.7 (AR(2)), n=4
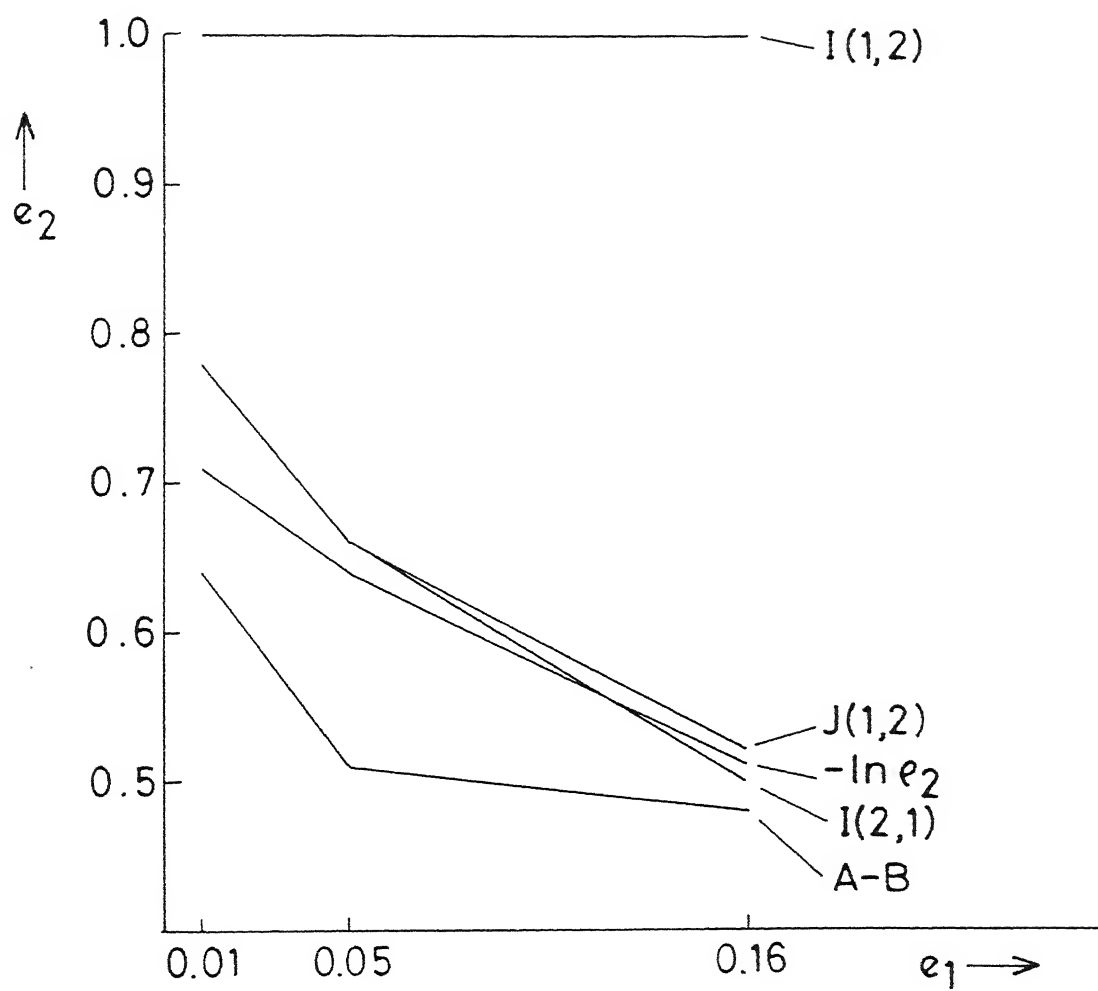
Fig.3.13 Example 3.3.7 (AR (2)). n = 5
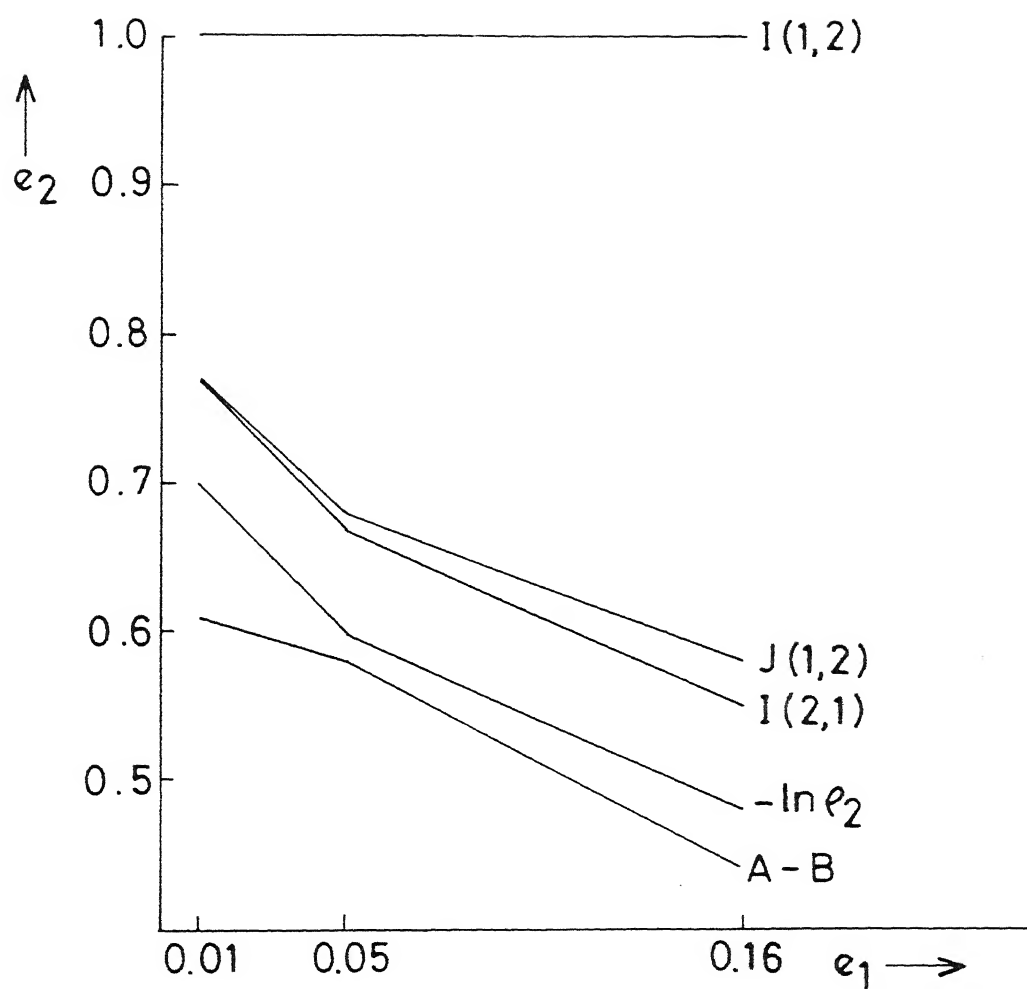
Fig.3.14 Example 3.3.7 (AR (2)), n = 10
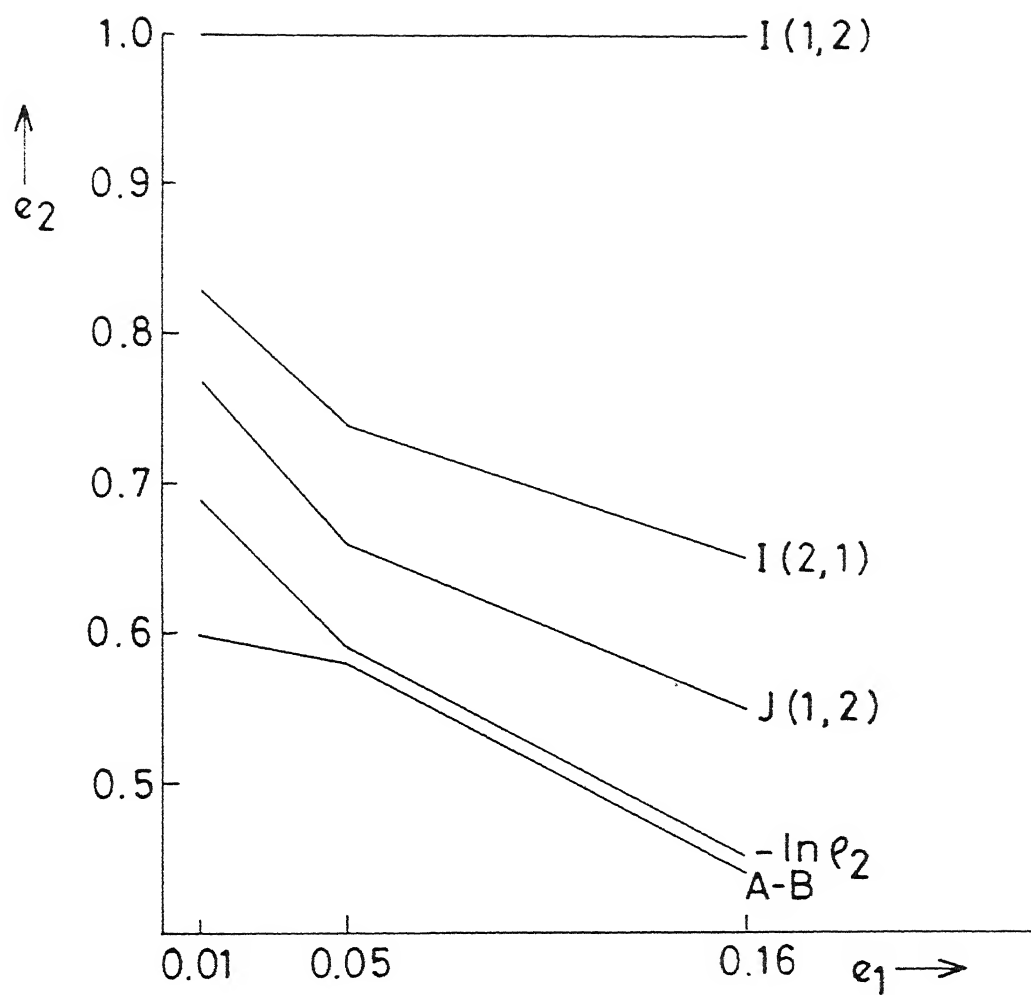
Fig. 3.15 Example 3.3.7 (AR (2)). n = 20
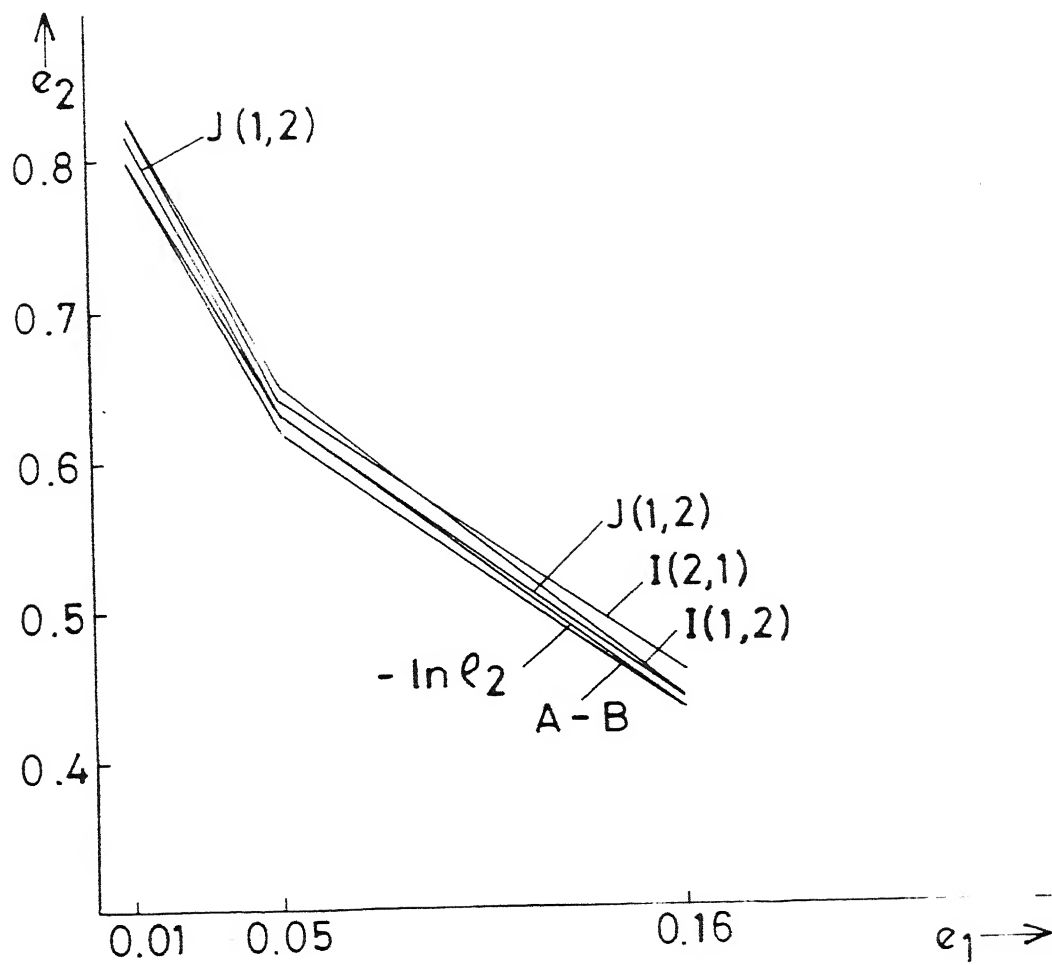
Fig. 3.16   Example 3.3.8 (AR(2)), n = 2
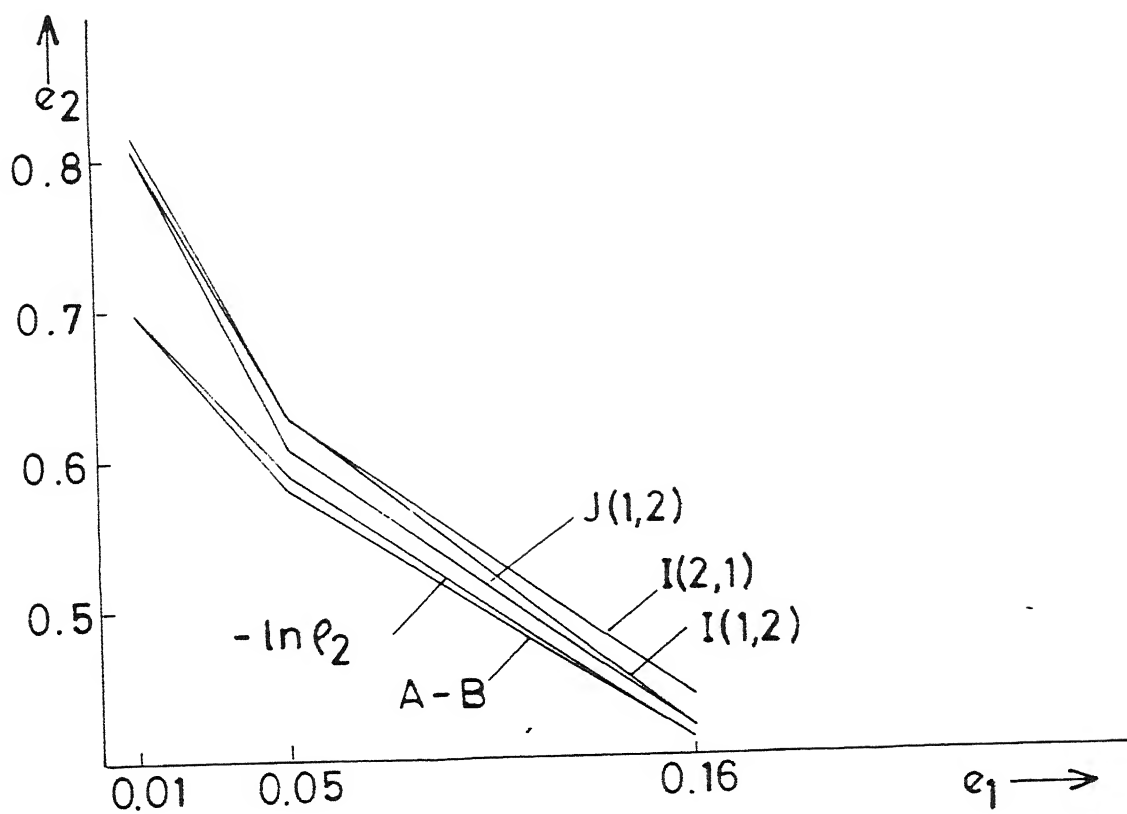
Fig. 3.17   Example 3.3.8 (AR(2)), n = 20
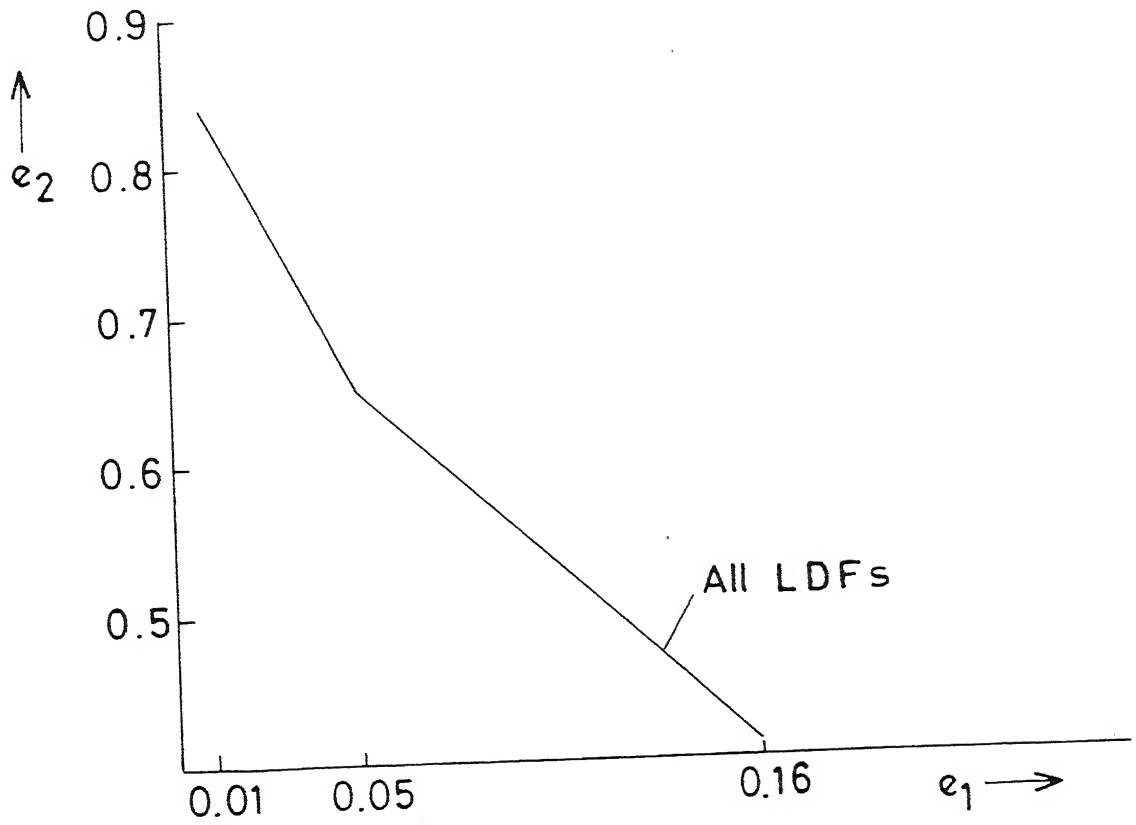
Fig.3.18 Example 3.3.9 (AR(2)), n = 2
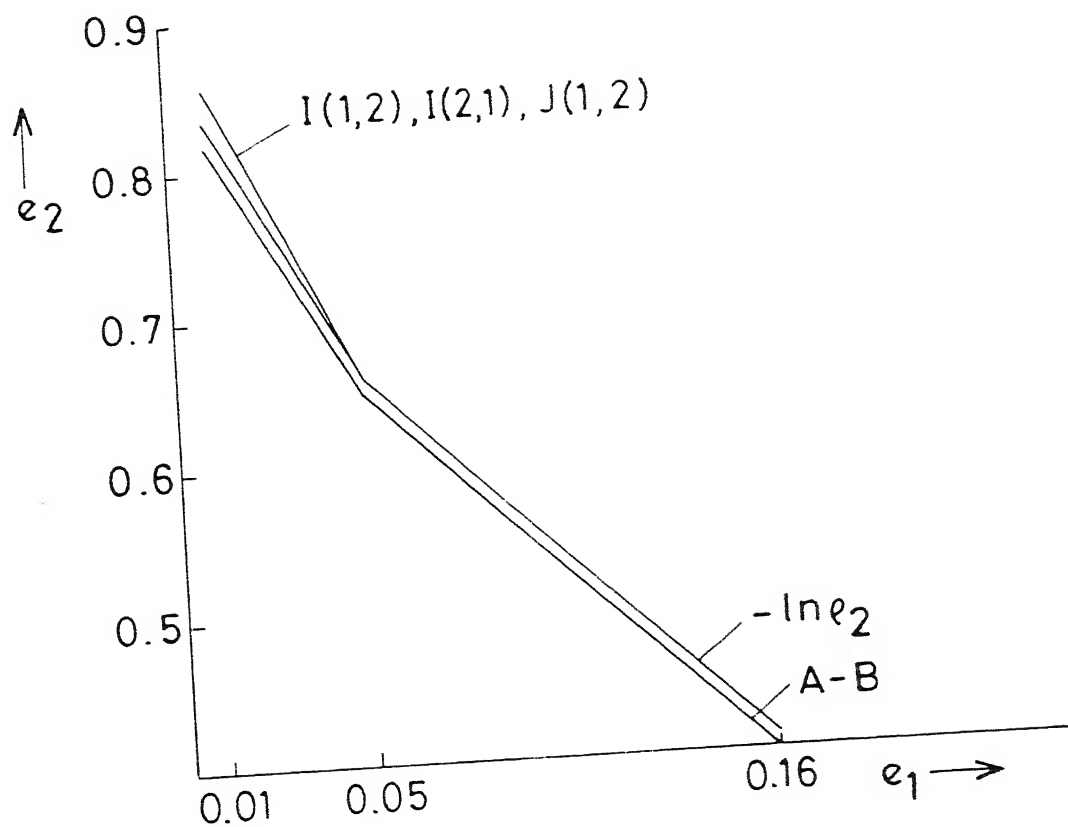
Fig .3.19 Example 3.3.9 (AR (2)). n = 20

Table 3.8

Results due to $-\ln\rho_2(1,2;y)$

Example	$\hat{\theta}$	itar	n	$e_1=.01$	$e_1=.05$	$e_1=.16$
	−1.0000	1	2	0.9082	0.7389	0.5000
	0.1389	9	3	0.7157	0.6141	0.5120
	0.1152	10	4	0.7190	0.6331	0.5120
3.3.7	0.1943	11	5	0.7734	0.6480	0.5120
	0.1760	11	10	0.7000	0.6000	0.4800
	0.1539	10	20	0.6900	0.5900	0.4500
	−0.1416	5	2	0.8051	0.6293	0.4325
	−0.1417	5	3	0.8051	0.6293	0.4325
	−0.1510	4	4	0.8051	0.6293	0.4325
3.3.8	−0.1512	3	5	0.8051	0.6293	0.4325
	−0.1654	3	10	0.7831	0.6103	0.4209
	−0.1720	3	20	0.7054	0.5980	0.4116
	−0.8802	2	2	0.8413	0.6591	0.4129
	−0.8816	2	3	0.8413	0.6591	0.4129
	−0.8817	2	4	0.8413	0.6591	0.4129
3.3.9	−0.8816	2	5	0.8413	0.6591	0.4129
	−0.8819	2	10	0.8456	0.6580	0.4089
	−0.8881	2	20	0.8350	0.6572	0.4100

Table 3.9

Results due to $I(1,2;y)$

Example	$\hat{\theta}$	iter	n	$e_1=.01$	$e_1=.05$	$e_1=.16$
	•	•	2	•	•	•
	8.2614	8	3	1.0	1.0	1.0
	3.4601	10	4	1.0	1.0	1.0
	3.9963	5	5	1.0	1.0	1.0
3.3.7	4.6401	6	10	1.0	1.0	1.0
	5.4059	5	20	1.0	1.0	1.0
	−1.8069	11	2	0.8365	0.6554	0.4443
	−2.0192	14	3	0.8365	0.6554	0.4443
	−1.9102	10	4	0.8365	0.6554	0.4443
3.3.8	−2.3411	9	5	0.8365	0.6554	0.4443
	−2.4500	6	10	0.8229	0.6417	0.4332
	−1.1999	5	20	0.8100	0.6339	0.4200
	−29.2505	3	2	0.8413	0.6591	0.4129
	−30.1821	4	3	0.8413	0.6591	0.4129
	−30.2233	4	4	0.8413	0.6591	0.4129
3.3.9	−30.3009	4	5	0.8414	0.6591	0.4129
	−30.4103	4	10	0.8414	0.6591	0.4129
	−30.4350	4	20	0.8591	0.6591	0.4129

Table 3.10

Results due to I(2,1;y)

Example	$\hat{\theta}$	iter	n	$e_1=.01$	$e_2=.05$	$e_1=.16$
	0.1829	4	2	0.7324	0.6026	0.4602
	0.1829	4	3	0.7357	0.7324	0.5000
3.3.7	0.2622	5	4	0.8159	0.6484	0.5000
	0.2157	5	5	0.7764	0.6554	0.5000
	0.1783	5	10	0.7700	0.6700	0.5500
	0.1661	3	20	0.8300	0.7400	0.6500
	0.1913	3	2	0.8361	0.6413	0.4637
	0.1605	4	3	0.8361	0.6413	0.4637
3.3.8	0.1707	4	4	0.8361	0.6413	0.4637
	0.1812	6	5	0.8361	0.6413	0.4637
	0.2001	8	10	0.8301	0.6356	0.4501
	0.2239	10	20	0.8229	0.6300	0.4400
	0.0296	2	2	0.8413	0.6591	0.4129
	0.0297	2	3	0.8413	0.6591	0.4129
3.3.9	0.0296	2	4	0.8413	0.6591	0.4129
	0.0296	2	5	0.8413	0.6591	0.4129
	0.0203	2	10	0.8571	0.6582	0.4091
	0.0296	2	20	0.8591	0.6591	0.4120

Table 3.11

Results due to $J(1,2;y)$

Example	$\hat{\theta}$	iter	n	$e_1=.01$	$e_1=.05$	$e_1=.16$
	-1.0000	1	2	0.9099	0.7422	0.5000
	0.1418	7	3	0.7190	0.6217	0.5160
	0.2470	10	4	0.8212	0.6484	0.5280
3.3.7	0.2106	7	5	0.7764	0.6517	0.5160
	0.1779	7	10	0.7700	0.6800	0.5800
	0.1539	8	20	0.7700	0.6600	0.5500
	0.0521	5	2	0.8215	0.6293	0.4443
	0.0434	5	3	0.8215	0.6293	0.4443
	0.0356	4	4	0.8215	0.6293	0.4443
3.3.8	0.0341	3	5	0.8215	0.6293	0.4443
	0.0200	3	10	0.8200	0.6201	0.4300
	0.0010	3	20	0.8100	0.6109	0.4205
	-0.8019	2	2	0.8413	0.6591	0.4129
	-0.8040	2	3	0.8413	0.6591	0.4129
	-0.8041	2	4	0.8413	0.6591	0.4129
3.3.9	-0.8041	2	5	0.8413	0.6591	0.4129
	-0.8142	2	10	0.8413	0.6591	0.4129
	-0.8152	2	20	0.8411	0.6591	0.4129

Table 3.12

Results due to the A–B procedure

Example	$\hat{\theta}$	iter	n	$e_1=.01$	$e_1=.05$	$e_1=.16$
	0.4647	6	2	0.7673	0.6406	0.4880
	0.2332	31	3	0.6293	0.5753	0.4880
3.3.7	0.1435	7	4	0.6133	0.5871	0.4840
	0.3952	7	5	0.6389	0.5120	0.4801
	−0.0947	14	10	0.6100	0.5800	0.4400
	−0.0792	14	20	0.6000	0.5800	0.4400
	0.1503	3	2	0.7995	0.6255	0.4325
	0.1621	5	3	0.7995	0.6255	0.4325
3.3.8	0.1654	5	4	0.7995	0.6255	0.4325
	0.1637	4	5	0.7995	0.6255	0.4325
	0.1801	4	10	0.7800	0.6019	0.4211
	0.1800	4	20	0.6992	0.5801	0.4116
	0.2461	2	2	0.8413	0.6591	0.4129
	0.2461	2	3	0.8413	0.6591	0.4129
3.3.9	0.2460	2	4	0.8413	0.6591	0.4129
	0.2461	2	5	0.8413	0.6591	0.4129
	−0.2106	2	10	0.8351	0.6562	0.4089
	−0.2203	2	20	0.8200	0.6500	0.4000

We observe the following from the above Tables.

(1) In Example 3.3.7 , the LDFs due to $-\ln\rho_2(1,2;y)$ do better than the LDFs obtained by maximizing other distances and the A-B procedure does little better than this. But the finding of the optimal θ in the A-B procedure requires more iterations. The maximization of $I(1,2;y)$ does not provide an optimal θ for n = 2, because during the iteration, θ becomes undefined due to the fact that $\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}$ becomes equal to $\underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha}$.

(2) The number of iterations taken in getting an optimal θ in Example 3.3.7 , is much higher than that in other Examples in general.

(3) In Example 3.3.8, and Example 3.3.9, all the LDFs have the same performance upto n = 5 and for n = 10, n = 20, we have the same conclusion as in Example 3.3.7.

From the above experiments (based on AR(1) and AR(2)) we can infer that the A-B procedure (which is admissible) does a little better than the LDFs depending on the Bhattacharyya distance in majority of the cases considered. And the LDFs yielded by Bhattacharyya distance do better than the LDFs given by other distances in almost all the cases. But since the A-B procedure suffers from computational difficulties (not only that for each e_1 value we have to find the $\underset{\sim}{\alpha}$ vector anew but also that it takes sometimes more computer - time to find it), the Bhattacharyya distance is preferrable. This

claim is strengthened as we shall see in later sections when we consider the covariance stationary time series wherein only our criterion of maximizing the Bhattacharyya distance admits an analytical solution for the required $\underset{\sim}{\alpha}$-vector asymptotically.

Remark 3.3.7 : The above conclusions do not go against the claim made earlier that our procedure, the procedures (due to maximizing other distances) and the A-B procedure are all admissible. To see that let us consider the Example 3.3.1 again. Having obtained the required $\underset{\sim}{\alpha}$'s by maximizing the distances we find c according to the following relation

$$c = \underset{\sim}{\delta}' R_\theta^{-1} \underset{\sim}{\mu}_1 - \underset{\sim}{\delta}' R_\theta^{-1} R_1 R_\theta^{-1} \underset{\sim}{\delta} , \qquad (3.3.43)$$

given in the Anderson-Bahadur theorem (see Remark 3.6.3); thus we can compute y_1 and y_2's . We find y_2 due to the A-B procedure for the given y_1 that results from the use of the LDF yielded by our criterion. The computations are shown in the Table 3.13 .

Table 3.13. (Admissibility in Example 3.3.1)

	$-\ln\rho_2(1,2;y)$	$I(1,2;y)$	$I(2,1;y)$	$J(1,2;y)$	A-B
y_1	1.5600	0.7265	5.1375	4.2715	1.5600
y_2	1.3082	1.7218	-0.4140	-0.1242	1.3082

It is clear from the above Table that our procedure is admissible and admissible procedures are not comparable. It may be pointed out that in the above AR(1) and AR(2) examples, c was not chosen according to (3.3.43).

Remark 3.3.8 : We have derived the covariance matrices for AR(1) and AR(2) by solving some difference equations. They can also be obtained by an interesting method which we shall describe now.

A pth order autoregressive scheme is defined as

$$Z(t+p) = \beta_1 Z(t+p-1) + \beta_2 Z(t+p-2) + \ldots + \beta_p Z(t) + \varepsilon(t+p),$$
$$(t = 0,1,\ldots,n-1,\ldots)$$

which can also be written as

$$
\begin{bmatrix} Z(t+1) \\ Z(t+2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Z(t+p) \end{bmatrix}
=
\begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ & & \cdot & & \cdot \\ & & \cdot & & \cdot \\ & & \cdot & & \\ 0 & 0 & 0 & \ldots & 1 \\ \beta_p & \beta_{p-1} & \ldots & \beta_1 \end{bmatrix}
\begin{bmatrix} Z(t) \\ Z(t+1) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Z(t+p-1) \end{bmatrix}
+
\begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \varepsilon(t+p) \end{bmatrix}
$$

$$(3.3.44)$$

Let

$$
\underset{\sim}{Y}_{t+p} =
\begin{bmatrix} Z(t+1) \\ Z(t+2) \\ \vdots \\ Z(t+p) \end{bmatrix}, \quad
\underset{\sim}{u}_{t+p} =
\begin{bmatrix} 0 \\ 0 \\ \vdots \\ \varepsilon(t+p) \end{bmatrix}
$$

Then (3.3.44) reduces to

$$\underset{\sim}{y}_{t+p} = B\underset{\sim}{y}_{t+p-1} + \underset{\sim}{u}_{t+p}$$

$$\Rightarrow \underset{\sim}{y}_n = \sum_{i=0}^{n-1} B^i \underset{\sim}{u}_{n-i} \text{ (on the assumption } y_o \equiv 0).$$

Then $H_{pxp} \overset{\Delta}{=} E\underset{\sim}{y}_i \underset{\sim}{y}'_j = E(\sum_{k=1}^{i-1} B^k \underset{\sim}{u}_{i-k})(\sum_{k=1}^{i-1} \underset{\sim}{u}'_{j-k} B'^k)$

$$= \sum_{\ell} \sum_{k} B^k (E\underset{\sim}{u}_{i-k} \underset{\sim}{u}'_{j-\ell}) B'^\ell$$

$$= (\sum_{k} B^k L B'^k) B'^{(j-i)} \text{ (where } E\underset{\sim}{u}_t \underset{\sim}{u}'_t = L = \begin{bmatrix} 0 & 0 \dots 0 \\ 0 & \dots 0 \\ \vdots \\ 0 & 0 \dots 1 \end{bmatrix})$$

$$= (\sum_{k=1}^{i-1} B_1^k B_1'^k) B'^{(j-i)} \text{ (where } B_1 = BL)$$

$$= [I-(B_1 B_1')^i][I-B_1 B_1']^{-1} (B')^{(j-i)} \qquad (3.3.45)$$

Hence $EZ(i)Z(j) = (i,j)$th element of the required covariance matrix $= (p,p)$th element of H.

Remark 3.3.9 : The general form of our linear discriminant function (LDF) is

$$y_\theta \overset{\Delta}{=} \underset{\sim}{\delta}' R_\theta^{-1} \underset{\sim}{x}$$

It follows from (3.3.6).

Remark 3.3.10 : Noting the form of $\underset{\sim}{\alpha}$ in (3.3.6) we may restrict our attention to the class

$$C \overset{\Delta}{=} \{ \underset{\sim}{\alpha} : \underset{\sim}{\alpha} = R_\theta^{-1} \underset{\sim}{\delta} , \theta \text{ being \underline{any} scalar } \ni R_\theta \text{ is non-singu}$$

for an $\underset{\sim}{\alpha}$ that would maximize $-\ln\rho_2(1,2;y_\theta)$. Then the maximizing

condition is a polynomial in θ and the properties of the roots can be studied. We illustrate this for the Example 3.3.1.

We have

$$\rho_2(1,2;Y_\theta) = \frac{(\sigma_1^2\sigma_2^2)^{\frac{1}{4}}}{\{\frac{1}{2}(\sigma_1^2+\sigma_2^2)\}^{\frac{1}{2}}} \exp\{-\frac{1}{4}\frac{(\underline{\delta}'R_\theta^{-1}\underline{\delta})^2}{\sigma_1^2+\sigma_2^2}\}(\text{see } 2.2.3) \qquad (3.3.46)$$

where $\sigma_j^2 \overset{\Delta}{=} \underline{\delta}'R_\theta^{-1}R_jR_\theta^{-1}\underline{\delta}$

$$R_\theta = \begin{bmatrix} 6.92-36.75\theta & -5.27-13.92\theta \\ & \\ & 40.89-287.92\theta \end{bmatrix}$$

One can show easily

$$\underline{\delta}'R_\theta^{-1}\underline{\delta} = \frac{10792.23}{|R_\theta|}(0.4262-4.1789\theta)$$

$$\sigma_1^2 = \frac{(48176798)(2.4176)}{|R_\theta|^2}(\theta^2-0.198\theta+0.01),$$

and $$\sigma_2^2 = \frac{(48176798)(2.4176)}{|R_\theta|^2}(4.022\theta^2-0.8203\theta+0.045)$$

Then, after some simplifications, $\frac{\partial \ln \rho_2}{\partial \theta} = 0$ reduces to

$$-0.75\theta^6 + 0.07032\theta^5 + 0.08\theta^4 - 0.0242\theta^3 + 0.003\theta^2 - 0.0002\theta =$$

We have solved this polynomial by Graeffe's Root-Squaring method (Appendix D). The computer output is given below :

Table 3.14

Polynomial method in Example 3.3.1

Root	Value of the polynomial	Iter-ation	Conclusion	The sign of $\dfrac{\partial^2 \ln \rho_2}{\partial \rho^2}$
-0.4151910	-0.1208365E-10	5	possibly a root	> 0
0.204327	0.1619732E-04	5	possibly a root	< 0
0.000000	0.0000E+00	5	is a root	< 0
15.657150	0.14638E+08	5	possibly not a root	—

From Table 3.14 we conclude that $-\ln \rho_2(1,2;y_\theta)$ has a maximum at $\theta = -.415191$ and this is consistent with our earlier finding (see Section 3.3.1).

3.4 COMPARISON OF THE BEHAVIOUR OF THE LDF OBTAINED BY MAXIMIZING THE BHATTACHARYYA DISTANCE WITH THE QUADRATIC (OPTIMAL) DISCRIMINANT FUNCTION WHEN $R_2 = dR_1$

We consider the case when $R_2 = dR_1$, $d > 0$ scalar , as in ([19]).

Then the LDF obtained by maximizing the Bhattacharyya distance is given by $\dfrac{\delta' R_1^{-1} x}{d+1}$; since for $R_2 = dR_1$, (3.3.2)' becomes

$$\ln \frac{\{(\alpha' R_1 \alpha) d (\alpha' R_1 \alpha)\}^{\frac{1}{4}}}{\{\frac{1}{2}(d+1)\alpha' R_1 \alpha\}^{\frac{1}{2}}} - \frac{1}{4} \frac{(\alpha' \delta)^2}{\alpha'(R_1 + dR_1)\alpha} \quad , \text{ which in turn}$$

equals
$$\left[\ln\left(\frac{d^{\frac{1}{4}}}{\{\frac{1}{2}(d+1)\}^{\frac{1}{2}}}\right) - \frac{1}{4}\frac{(\underset{\sim}{\alpha}'\underset{\sim}{\delta})^2}{\underset{\sim}{\alpha}'[(d+1)R_1]\underset{\sim}{\alpha}}\right]$$

Thus our linear procedure is given by

$$\frac{\underset{\sim}{\delta}' R_1^{-1} \underset{\sim}{x}}{d+1} \overset{H_1}{\underset{H_2}{\gtrless}} c' \qquad (3.4.1)$$
$$(\text{Accept})$$

Apply the following transfomation :

$$\underset{\sim}{Y} = R_1^{-1/2}(\underset{\sim}{\mu}_1 - \underset{\sim}{X}) \qquad (3.4.2)$$

This reduces R_1 to the identity matrix and R_2 to dI.

Then, $Y \sim N_n(0,I)$ under H_1

$Y \sim N_n(\underset{\sim}{\nu}, dI)$ under H_2, where $\underset{\sim}{\nu} = R_1^{-1/2}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2) = R_1^{-1/2}\underset{\sim}{\delta}$

$$(3.4.3)$$

Now, $\frac{\underset{\sim}{\delta}' R_1^{-1}\underset{\sim}{x}}{d+1} = -\frac{\underset{\sim}{\nu}' R_1^{\frac{1}{2}} R_1^{-1}(\underset{\sim}{\mu}_1 - \underset{\sim}{X})}{d+1} + \frac{\underset{\sim}{\nu}' R_1^{-\frac{1}{2}}\underset{\sim}{\mu}_1}{d+1}.$

$$= \frac{-\underset{\sim}{\nu}'\underset{\sim}{Y}}{d+1} + c_1$$

Then (3.4.1) reduces to

$$\frac{\underset{\sim}{\nu}'\underset{\sim}{Y}}{d+1} \overset{H_1}{\underset{H_2}{\gtrless}} c \qquad (3.4.4)$$
$$(\text{reject})$$

The errors of misclassifications associated with (3.4.4) are given below :

$$e_1 = 1 - \Phi\left(\frac{c}{\left\{\frac{\nu'\nu}{(d+1)^2}\right\}^{\frac{1}{2}}}\right)$$

$$= 1 - \Phi\left(\frac{c(d+1)}{\left\{T^2[\omega+\overline{1-\omega}d]\right\}^{\frac{1}{2}}}\right) \tag{3.4.5}$$

where $\quad T^2 \overset{\Delta}{=} \dfrac{\nu'\nu}{\omega+\overline{1-\omega}d}$ $\tag{3.4.6}$

and ω is the prior probability of H_1.

$$e_2 = \Phi\left(\frac{c - \frac{\nu'\nu}{d+1}}{\left\{\frac{d}{(d+1)^2}\nu'\nu\right\}^{\frac{1}{2}}}\right)$$

$$= \Phi\left(\frac{\left[c - \frac{\nu'\nu}{\omega+\overline{1-\omega}\alpha}\frac{\omega+\overline{1-\omega}d}{d+1}\right](d+1)}{\left\{\frac{\nu'\nu}{\omega+\overline{1-\omega}d}\cdot\left[\omega+\overline{1-\omega}d\right]d\right\}^{\frac{1}{2}}}\right)$$

$$= \Phi\left(\frac{c(d+1) - T^2[\omega+\overline{1-\omega}d]}{\left\{dT^2[\omega+\overline{1-\omega}d]\right\}^{\frac{1}{2}}}\right) \tag{3.4.7}$$

Now the total probability of misclassification is defined as

$$\omega e_1 + (1-\omega)e_2 . \tag{3.4.8}$$

The cut-off point c is chosen to minimize (3.4.8).

The value of c for which (3.4.8) is a minimum satisfies the
equation (putting the first derivative = 0) :

$$\frac{\omega(d+1)}{(T^2[\omega + \overline{1-\omega d}])^{\frac{1}{2}}} \; e^{-\frac{1}{2} \; \frac{c^2(d+1)^2}{T^2[\omega + \overline{1-\omega d}]}}$$

$$= \frac{(1-\omega)(d+1)}{\{dT^2[\omega + \overline{1-\omega d}]\}^{\frac{1}{2}}} \; e^{-\frac{1}{2} \; \frac{\{c(d+1)-T^2[\omega + \overline{1-\omega d}]\}^2}{dT^2[\omega + \overline{1-\omega d}]}}$$

$$(3.4.9)$$

$$\text{or } \ln(\frac{\omega}{1-\omega} \; (d)^{\frac{1}{2}}) = -\frac{1}{2} \; \frac{\{c^2(d+1)^2 + T^4[\omega + \overline{1-\omega d}]^2 - 2cT^2\{d+1\}[\omega + \overline{1-\omega d}]\}}{dT^2[\omega + \overline{1-\omega d}]}$$

$$+ \frac{1}{2} \; \frac{c^2(d+1)^2}{T^2[\omega + \overline{1-\omega d}]}$$

$$= -\frac{1}{2} \; \frac{(d+1)^2}{dT^2[\omega + \overline{1-\omega d}]} \; [c^2 - dc^2 + T^4(\frac{\omega + \overline{1-\omega d}}{d+1})^2$$

$$- 2c \; T^2 \; \frac{\omega + \overline{1-\omega d}}{d+1}]$$

$$\text{or, } (1-d)c^2 - 2cT^2 \; \frac{\omega + \overline{1-\omega d}}{d+1} + T^4(\frac{\omega + \overline{1-\omega d}}{d+1})^2$$

$$+ \frac{2dT^2[\omega + \overline{1-\omega d}]}{(d+1)^2} \; \ln \frac{\omega}{1-\omega} \; (d)^{1/2} = 0 \qquad (3.4.10)$$

That is, the required c satisfies a quadratic equation, solutions

of which are given by (after some simplifications),

$$c = \frac{[\omega + \overline{1-\omega}d]}{(d+1)(1-d)} [T^2 \pm (dT^2)^{\frac{1}{2}} \{T^2 + \frac{d-1}{\omega + \overline{1-\omega}d}(\ln d + 2\ln \frac{\omega}{1-\omega})\}^{\frac{1}{2}}],$$

$$\text{when } d \neq 1. \qquad (3.4.11)$$

and $c = \frac{1}{4}T^2 + \frac{1}{2}\ln \frac{\omega}{1-\omega}$ when $d = 1$ (cf. (3.4.10)) $\qquad$ (3.4.12)

Remark 3.4.1 : The only instance in which the optimal cutpoint is not given by (3.4.11) is when assigning every observation to the same population yields a lower total probability of misclassification. This probability will be $\min(\omega, 1-\omega)$ and can be obtained by making c infinite.

Remark 3.4.2 : Once we have found c via (3.4.11), our desired test (3.4.4) is completely specified.

Remark 3.4.3 : The root corresponding to the negative sign in (3.4.11) gives the required minimum. This can be shown by computing the second derivative of the total probability of misclassification.

$$\frac{\partial^2}{\partial c^2} [\omega e_1 + \overline{1-\omega}e_2] =$$

$$= \frac{(d+1)\omega}{(T^2[\omega+\overline{1-\omega}d])^{\frac{1}{2}}} \{\frac{(d+1)^2 c}{T^2[\omega + \overline{1-\omega}d]}\} e^{-\frac{1}{2}\frac{c^2(d+1)^2}{T^2[\omega + \overline{1-\omega}d]}}$$

$$- \frac{(1-\omega)(d+1)^2}{(dT^2[\omega+\overline{1-\omega}d])^{\frac{1}{2}}} \{\frac{(d+1)c-T^2[\omega+\overline{1-\omega}d]}{dT^2[\omega + \overline{1-\omega}d]}\} e^{-\frac{1}{2}\frac{(c(d+1)-T^2[\omega+\overline{1-\omega}d])}{dT^2[\omega + \overline{1-\omega}d]}}$$

$$= \frac{c(d+1)^2}{T^2[\omega + \overline{1-\omega}d]} \left\{ \frac{\omega(d+1)}{(T^2[\omega + \overline{1-\omega}d])^{\frac{1}{2}}} \; e^{-\frac{1}{2} \frac{c^2(d+1)^2}{T^2[\omega + \overline{1-\omega}d]}} \right. -$$

$$- \frac{(1-\omega)(d+1)}{(dT^2[\omega + \overline{1-\omega}d])^{\frac{1}{2}}} \; e^{-\frac{1}{2} \frac{\{c(d+1)-T^2[\omega + \overline{1-\omega}d]\}^2}{dT^2[\omega + \overline{1-\omega}d]}} \right\}$$

$$+ \frac{c(d+1)^3(1-\omega)}{T^2[\omega + \overline{1-\omega}d](dT^2[\omega + \overline{1-\omega}d])^{\frac{1}{2}}} \; e^{-\frac{1}{2} \frac{\{c(d+1)-T^2[\omega + \overline{1-\omega}d]\}^2}{dT^2[\omega + \overline{1-\omega}d]}}$$

$$- \frac{(1-\omega)(d+1)^2}{(dT^2[\omega + \overline{1-\omega}d])^{\frac{1}{2}}} \left\{ \frac{(d+1)c - T^2[\omega + \overline{1-\omega}d]}{dT^2[\omega + \overline{1-\omega}d]} \right\} \times$$

$$e^{-\frac{1}{2} \frac{\{c(d+1)-T^2[\omega + \overline{1-\omega}d]\}^2}{dT^2[\omega + \overline{1-\omega}d]}}$$

Using (3.4.9), we have

$$\frac{\partial^2[\omega e_1 + \overline{1-\omega}e_2]}{\partial c^2}$$

$$= \frac{(1-\omega)(d+1)^2}{T^2[\omega + \overline{1-\omega}d](dT^2[\omega + \overline{1-\omega}d])^{\frac{1}{2}}} \left\{ c(d+1) - \frac{c(d+1)}{d} + \frac{T^2[\omega + \overline{1-\omega}d]}{d} \right\}$$

$$\times \; e^{-\frac{1}{2} \frac{\{c(d+1) - T^2[\omega + \overline{1-\omega}d]\}^2}{dT^2[\omega + \overline{1-\omega}d]}}$$

$$> 0$$

iff $\quad c(d+1) - \dfrac{c(d+1)}{d} + \dfrac{T^2[\omega + \overline{1-\omega}d]}{d} > 0$

This happens if

$$c = \begin{cases} \dfrac{[\omega + \overline{1-\omega}d]}{(1-d^2)}[T^2-(dT^2)^{\frac{1}{2}}\{T^2 + \dfrac{d-1}{\omega+\overline{1-\omega}d}(\ln d + 2\ln\dfrac{\omega}{1-\omega})\}^{\frac{1}{2}}], \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } d < 1 \\[2em] \dfrac{[\omega + \overline{1-\omega}d]}{(1-d^2)}[T^2+ (dT)^{\frac{1}{2}}\{T^2 + \dfrac{d-1}{\omega+\overline{1-\omega}d}(\ln d + 2\ln\dfrac{\omega}{1-\omega})\}^{\frac{1}{2}}], \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } d > 1 \end{cases}$$

We shall now apply the same transformation as (3.4.2)
to the quadratic discriminant function (QDF) described in
Section 3.3. Then (3.2.1) reduces to ([19]),

$$Z = \sum_{i=1}^{n} Z_i^2 > K ,$$

where $\qquad Z_i^2 = \dfrac{|d-1|}{2d}(Y_i + \dfrac{\nu_i}{d-1})^2$

and $\qquad K = \ln\dfrac{\omega}{1-\omega} + \dfrac{n}{2}\ln d + \dfrac{[\omega + \overline{1-\omega}d]T^2}{2(d-1)}$

Since the Z_i are normal random variables, by a result of
Patnaik ([45]), the distribution of Z under $H_j(j = 1,2)$ can
be approximated by a multiple, c_j , of a central χ^2-distribution
with degrees of freedom ℓ_j , where c_j and ℓ_j are chosen
to satisfy

$$s_j \overset{\Delta}{=} E_{H_j}(Z) = E_{H_j}(c_j \chi^2(\ell_j)) = c_j\ell_j$$

and $\qquad v_j^2 \overset{\Delta}{=} \text{Var}_{H_j}(Z) = \text{Var}(c_j \chi^2(\ell_j)) = 2c_j^2 \ell_j \quad (j = 1,2).$

It is easily shown that ([19]),

$$s_1 = \frac{1}{2d}\{ \frac{[\omega + \overline{1-\omega}\, d\,]T^2}{|d-1|} + n\,|d-1|\}$$

$$s_2 = \frac{1}{2} \{ \frac{d[\omega + \overline{1-\omega}\, d\,]T^2}{|d-1|} + n\,|d-1|\}$$

$$v_1^2 = \frac{1}{d^2} \{ [\omega + \overline{1-\omega}d\,]T^2 + \frac{n(d-1)^2}{2}\}$$

$$v_2^2 = d[\omega + \overline{1-\omega}\, d\,]T^2 + \frac{n(d-1)^2}{2}$$

(3.4.13)

Thus e_1 and e_2 are approximated by

$$e_1 = P(\text{assign } \underset{\sim}{x} \varepsilon\ H_2/H_1 \text{ is true})$$
$$\doteq P(\chi^2(\ell_1) > \frac{K}{c_1})$$

(3.4.14)

and

$$e_2 = P(\text{assign } \underset{\sim}{x}\ H_1/H_2 \text{ is true})$$
$$\doteq P(\chi^2(\ell_2) < \frac{K}{c_2})$$

(3.4.15)

3.4.1 NUMERICAL RESULTS

For all combinations of the following parameter values :
$T^2 = 0,1,2,4,8$; $d = 0.1,0.2,0.5,1,2,5,10; \omega = \frac{1}{2}, \frac{2}{3}, \frac{5}{6}$ and
$n = 1,2,6,10$; the total probability of misclassification
resulting from the use of LDF considered here was calculated
using equations (3.4.5) and (3.4.7) and for the QDF using the
approximations (3.4.14) and (3.4.15). Some results are shown
graphically in the figures : Fig. 3.20, Fig. 3.21 and Fig. 3.22.
The numerical results are shown in the following Tables:
Table 3.15, Table 3.16, Table 3.17.
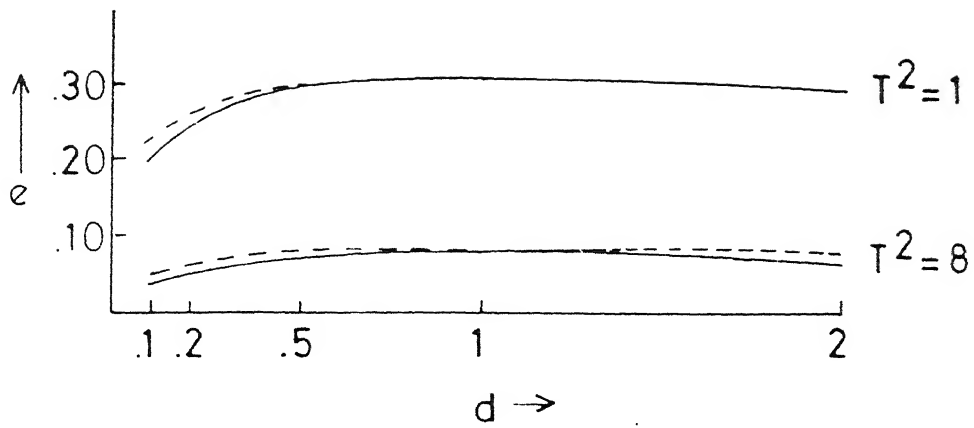
Fig. 3.20

Cont.

Fig. 3.20

$T^2 = 1$

$T^2 = 8$

$d \longrightarrow$

$n = 1$

--- Linear discriminant function

—— Quadratic discriminant function

$w = .6667$

$T^2 = 1$

$T^2 = 8$

$d \longrightarrow$

$n = 2$

Fig. 3.21

Cont.

Fig. 3.21

n = 1

--- Linear discriminant function

—— Quadratic discriminant function

w = .8333

n = 2

Fig. 3.22

.30

.20

e .10

.1 .2   .5   1   2

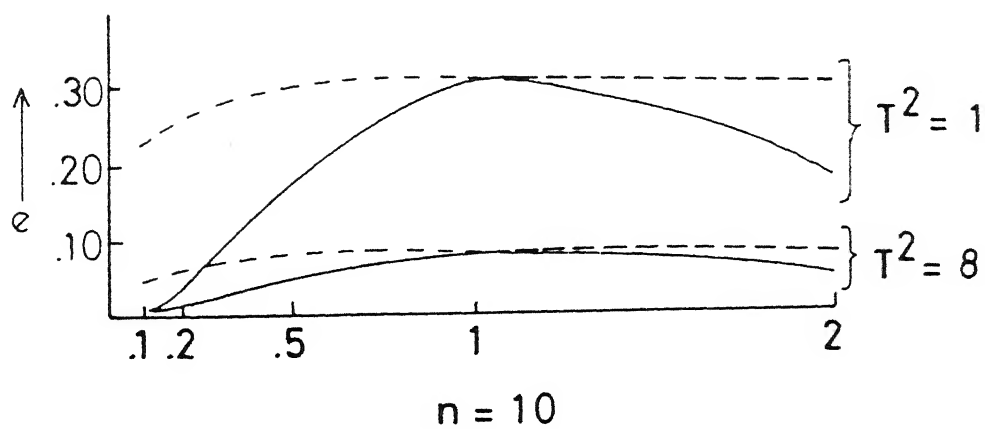$T^2 = 1$

$T^2 = 8$

n = 6

--- Linear discriminant function

—— Quadratic discriminant function

w = .8333

.30

.20

e .10

.1 .2   .5   1   2

$T^2 = 1$

$T^2 = 8$

n = 10
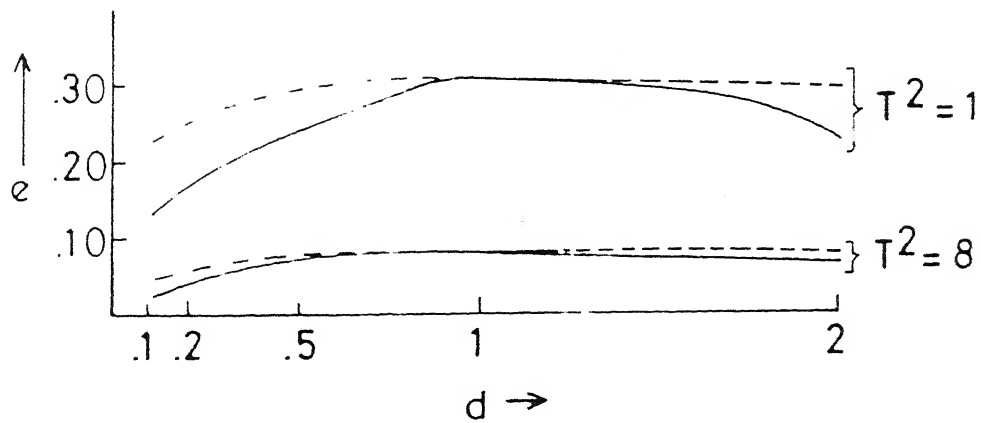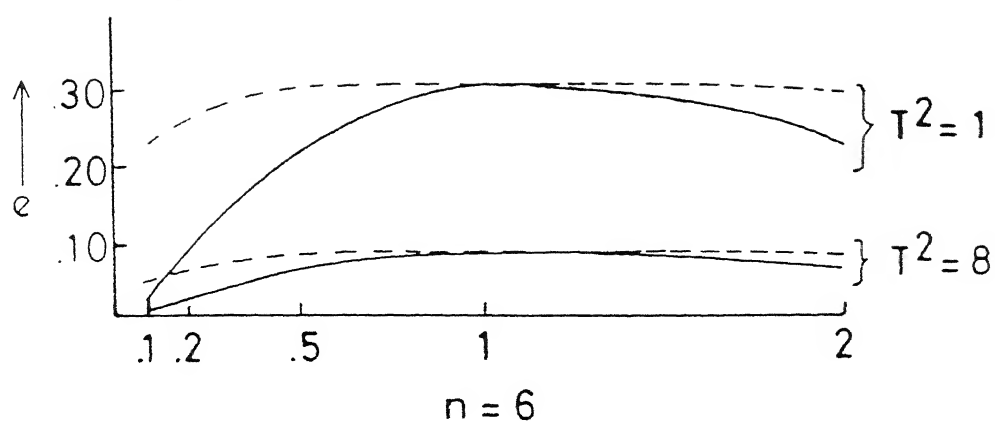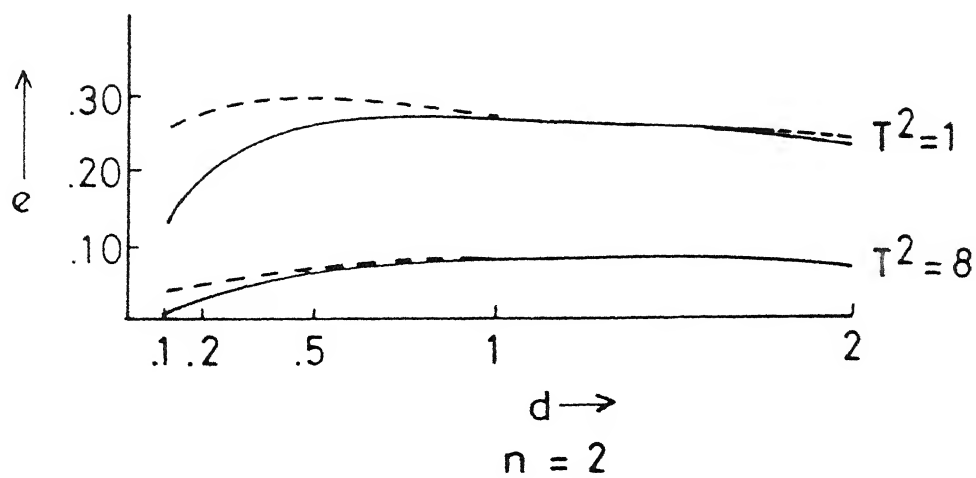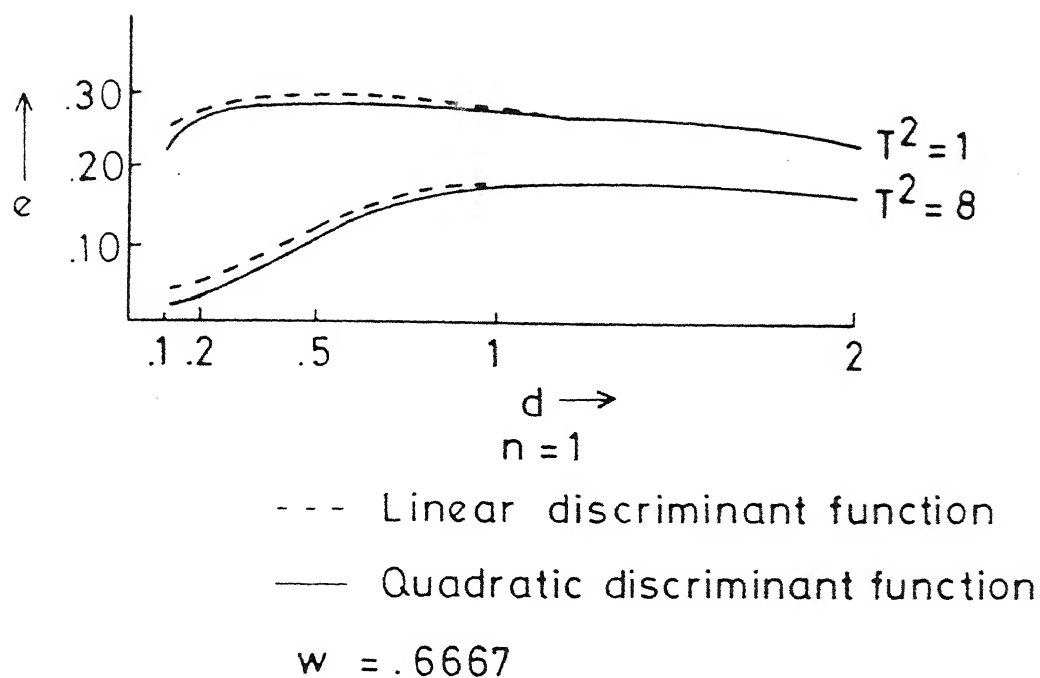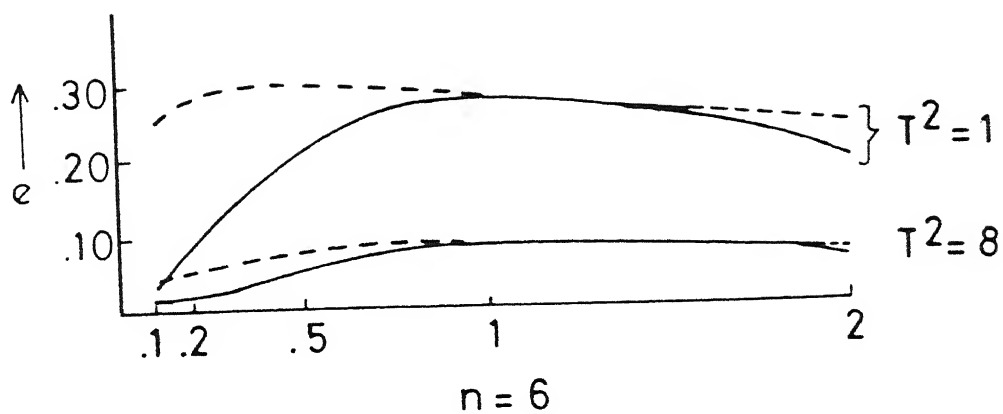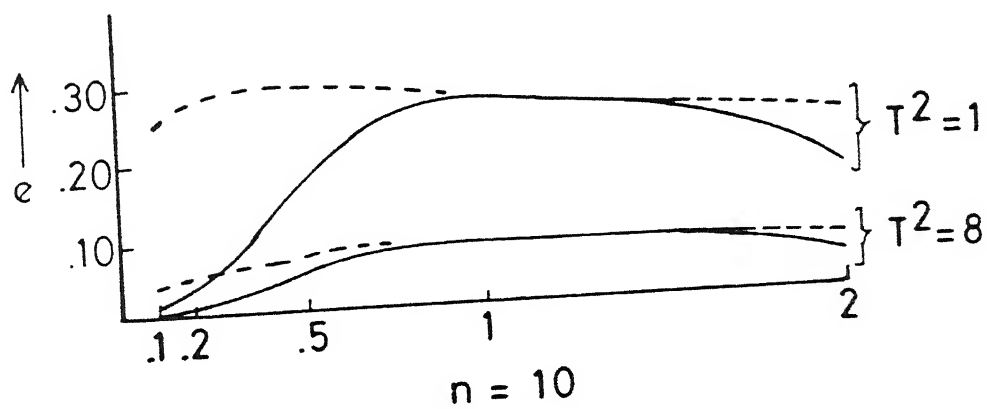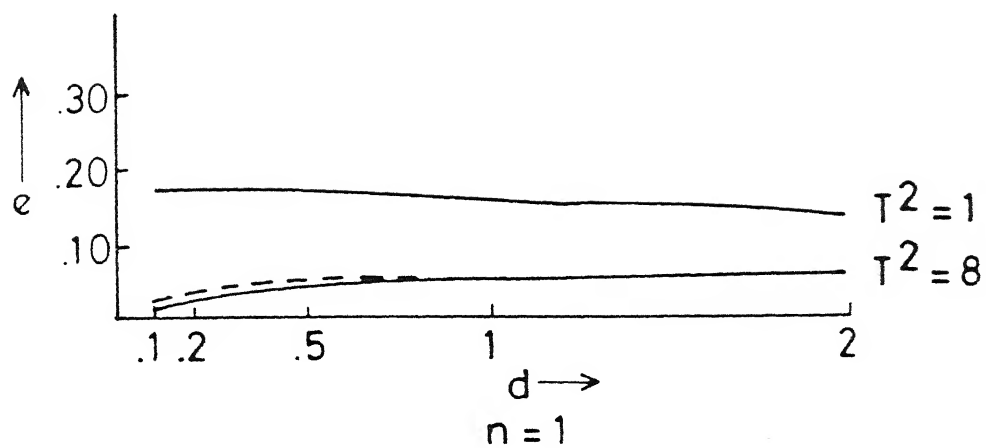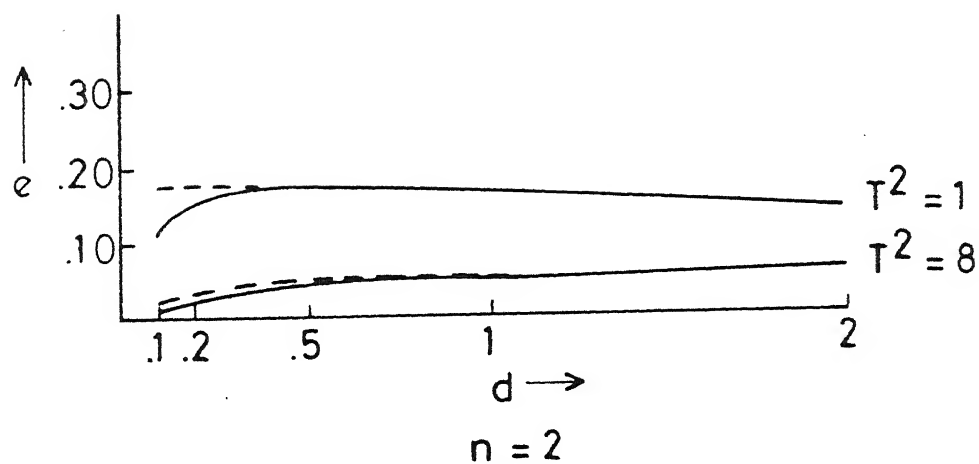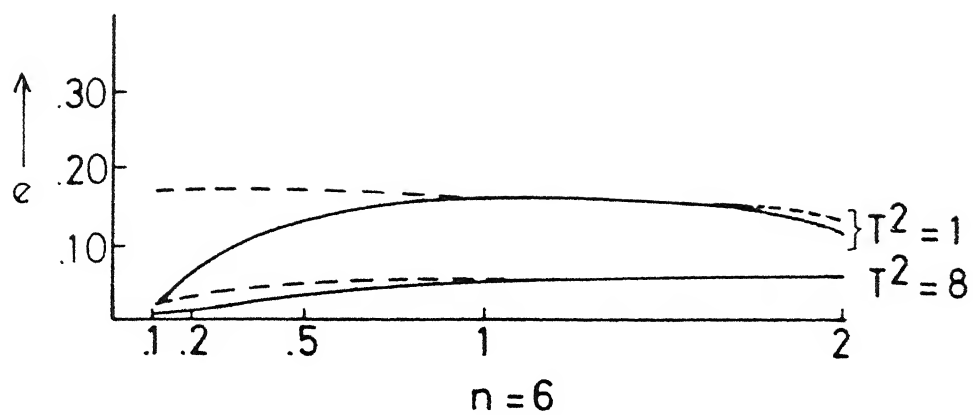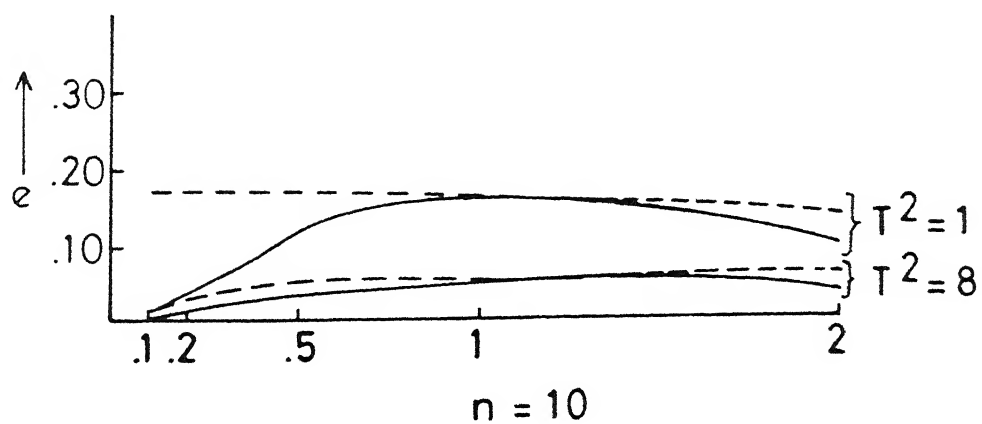
Fig. 3.22

Table 3.15

The probability of misclassification resulting from the use of LDF based on Bhattacharyya distance (for all n).

ω	T^2 d...	.1	.2	.5	1	2	5	10
.5000	0	.50	.50	.50	.50	.50	.50	.50
	1	.23	.26	.30	.31	.30	.26	.23
	2	.18	.20	.23	.24	.23	.20	.18
	4	.11	.13	.15	.16	.15	.13	.11
	8	.05	.06	.08	.08	.08	.06	.05
.6667	0	.33	.33	.33	.33	.33	.33	.33
	1	.26	.28	.29	.27	.24	.21	.18
	2	.18	.20	.22	.22	.20	.17	.15
	4	.10	.12	.14	.15	.14	.12	.10
	8	.04	.05	.07	.08	.07	.06	.06
.8333	0	.17	.17	.17	.17	.17	.17	.17
	1	.17	.17	.17	.16	.14	.12	.11
	2	.16	.16	.15	.14	.12	.11	.09
	4	.08	.09	.10	.10	.10	.08	.07
	8	.02	.03	.05	.05	.06	.05	.05

Table 3.16

The probability of misclassification obtained by using QDF (see [19]).

ω	T^2		$n = 1$								$n = 2$					
	d....	.1	.2	.5	1	2	5	10	.1	.2	.5	1	2	5	10
.5000	0	.25	.32	.42	.50	.42	.32	.25	.15	.24	.38	.50	.38	.24	.15
	1	.20	.25	.30	.31	.30	.25	.20	.13	.19	.23	.31	.23	.19	.13
	2	.17	.19	.23	.24	.23	.19	.17	.11	.15	.22	.24	.22	.15	.11
	4	.09	.12	.15	.16	.15	.12	.09	.07	.10	.14	.16	.14	.10	.07
	8	.04	.05	.07	.08	.07	.05	.04	.02	.04	.07	.08	.07	.04	.02
.6667	0	.27	.32	.33	.33	.31	.24	.18	.16	.24	.33	.33	.29	.19	.12
	1	.22	.27	.27	.27	.24	.20	.17	.13	.19	.26	.27	.23	.16	.09
	2	.16	.17	.20	.22	.20	.17	.14	.10	.15	.20	.22	.19	.13	.11
	4	.07	.09	.13	.15	.14	.11	.10	.06	.08	.12	.15	.14	.09	.07
	8	.02	.03	.06	.08	.07	.06	.04	.01	.03	.06	.08	.07	.05	.03
.8333	0	.17	.17	.17	.17	.16	.13	.10	.14	.17	.17	.17	.16	.11	.07
	1	.17	.17	.17	.16	.14	.12	.10	.11	.15	.17	.16	.14	.10	.06
	2	.12	.12	.13	.14	.12	.11	.09	.08	.11	.13	.14	.12	.09	.06
	4	.05	.07	.08	.10	.10	.08	.07	.04	.05	.09	.10	.09	.07	.05
	8	.01	.02	.04	.05	.06	.05	.05	.01	.02	.04	.05	.06	.05	.03

Table 3.17

The probability of misclassification obtained by using QDF (see [19]).

ω	T^2	d...		n = 6								n = 10					
		.1	.2	.5	1	2	5	10		.1	.2	.5	1	2	5	10
.5000	0	.03	.09	.28	.50	.28	.09	.03		.01	.04	.24	.50	.24	.04	.01
	1	.03	.08	.23	.31	.23	.08	.03		.01	.03	.18	.31	.18	.03	.01
	2	.02	.07	.18	.24	.18	.07	.02		.01	.03	.15	.24	.15	.03	.01
	4	.02	.04	.12	.16	.12	.04	.02		.01	.02	.10	.16	.10	.02	.01
	8	.01	.02	.06	.08	.06	.02	.01		.01	.01	.05	.08	.05	.01	.01
.6667	0	.03	.09	.27	.33	.23	.08	.02		.01	.04	.22	.33	.19	.04	.000
	1	.03	.08	.21	.27	.19	.07	.02		.01	.03	.17	.27	.16	.03	.001
	2	.02	.07	.17	.22	.16	.06	.02		.01	.03	.14	.22	.13	.02	.001
	4	.01	.04	.11	.15	.11	.05	.02		.00	.02	.09	.15	.10	.02	.001
	8	.01	.01	.05	.08	.06	.02	.01		.00	.01	.05	.08	.05	.01	.001
.8333	0	.03	.08	.16	.17	.14	.05	.01		.01	.04	.15	.17	.12	.02	.000
	1	.02	.06	.14	.16	.12	.04	.01		.01	.03	.12	.16	.10	.02	.001
	2	.02	.05	.12	.14	.11	.05	.01		.00	.03	.10	.14	.09	.02	.001
	4	.01	.03	.07	.10	.08	.03	.01		.00	.01	.07	.10	.07	.01	.001
	8	.01	.01	.03	.05	.05	.02	.01		.00	.01	.03	.05	.04	.01	.001

From the Tables , we can conclude the following :

1) For n = 1, QDF does little better than LDF

2) Agreement is adequate in the cases given below

(irrespective of the value of n) :

   i) for large T^2 value

   ii)for ω near 1

   iii)for moderate range of d values

3) Agreement becomes worse as n increases for a given

ω. This occurs for the obvious reason that with large

n, there are more variables with variance discrepencies

to be utilized.

3.5 TWO CLASSES OF TESTS (MODIFIED MINIMAX RULE)

As shown earlier, maximization of the Bhattacharyya

distance yields an $\underset{\sim}{\alpha}$ vector for the test given below

$$\underset{\sim}{\alpha}' \underset{\sim}{x} \underset{H_2}{\overset{H_1}{\gtrless}} c \qquad (3.5.1)$$

and a class of tests can be generated by assigning various

values to c.

Let us now consider an another method of specifying

the test (3.5.1). e_1 and e_2 are reproduced below for

convenience :

$$e_1 = 1 - \Phi(Y_1), \text{ where } Y_1 = \frac{\underset{\sim}{g}' \underset{\sim}{\mu} - c}{(\underset{\sim}{\alpha}' R_1 \underset{\sim}{\alpha})^{\frac{1}{2}}} \qquad (3.5.2)$$

and $\quad e_2 = 1 - \Phi(y_2)$, where $y_2 = \dfrac{c}{(\underset{\sim}{\alpha}' R_2 \underset{\sim}{\alpha})^{\frac{1}{2}}}$ $\qquad$ (3.5.3)

We have taken $\underset{\sim}{\mu}_2 = 0$ for simplicity and denote $\underset{\sim}{\mu}_1$ by $\underset{\sim}{\mu}$.
We propose the following criterion :

$$\min_{\substack{Pr(\varepsilon) \\ e_1 = k_1 e_2}} \quad \text{, for a given } k_1, \qquad (3.5.4)$$

or $\qquad \min_{\substack{e_2 \\ e_1 = k_1 e_2}}$ $\qquad\qquad\qquad\qquad$ (3.5.5)

or $\qquad$ equivalently, $\quad \max_{\substack{y_2 \\ y_1 = k_2 y_2}}$, for some k_2 $\qquad$ (3.5.6)

After some simplification, (3.5.6) reduces to

$$\max_{\underset{\sim}{\alpha}} \left\{ \frac{\underset{\sim}{\alpha}' \underset{\sim}{\mu}}{k_2 (\underset{\sim}{\alpha}' R_1 \underset{\sim}{\alpha})^{\frac{1}{2}} + (\underset{\sim}{\alpha}' R_2 \underset{\sim}{\alpha})^{\frac{1}{2}}} \right\} \qquad (3.5.7)$$

and this gives rise to another class of tests generated by k_2, for $\underset{\sim}{\alpha}$ and c are provided by this criterion. Now we shall establish a one-to-one correspondence between the two classes of tests described above which is embodied in the following theorem.

Before stating the theorem we need to define the notion of equivalence of two tests.

Definition 3.5.1 : A test completely specified by the pair $(\underset{\sim}{\alpha}, c)$ is said to be equivalent to another if each probability of misclassification of the former is equal to the corresponding one of the latter.

<u>Theorem 3.5.1</u> : Let (α_B, c_B) denote a test obtained by maximizing the Bhattacharyya distance and (α_{k_2}, c_{k_2}) denote that due to (3.5.7). Then

(a) for a given c_B , we can find a k_2 such that (α_B, c_B) is equivalent to (α_{k_2}, c_{k_2}) and $y_1 = k_2 y_2$ for both tests ; conversely, for a given k_2 and associated (α_{k_2}, c_{k_2}), there exists a c_B such that (α_{k_2}, c_{k_2}) is equivalent to (α_B, c_B) and $y_1 = k_2 y_2$ is maintained . The respective k_2 and c_B are given by the following relation :

$$c_B = \frac{(\alpha_B' \mu)\left(\dfrac{\alpha_B' R_2 \alpha_B}{\alpha_B' R_1 \alpha_B}\right)^{\frac{1}{2}}}{k_2 + \left(\dfrac{\alpha_B' R_2 \alpha_B}{\alpha_B' R_1 \alpha_B}\right)^{\frac{1}{2}}} , \qquad (3.5.8)$$

(b) there exists a k_2 given by

$$k_2 = -\frac{1}{\theta}\left(\frac{\alpha_B' R_1 \alpha_B}{\alpha_B' R_2 \alpha_B}\right)^{\frac{1}{2}} \qquad (3.5.9)$$

where $\qquad \alpha_B = (R_1 - \theta R_2)^{-1}\mu \qquad (3.5.10)$

such that $\qquad \alpha_{k_2} = \alpha_B \qquad$ and we obtain equivalent test by choosing $c_B = c_{k_2}$.

<u>Proof</u> : We see that the value of α for which the quantity within braces in (3.5.7) is a maximum satisfies

$$\alpha_{k_2} = \left(R_1 - \frac{\theta_1}{k_2} R_2\right)^{-1}\mu \qquad (3.5.11)$$

where
$$-\theta_1 = \left(\frac{\alpha'_{k_2} R_1 \alpha_{k_2}}{\alpha'_{k_2} R_2 \alpha_{k_2}}\right)^{\frac{1}{2}}$$
(3.5.12)

which follows by usual calculus procedures.

(a) Suppose, we are given k_2. Then (α_{k_2}, c_{k_2}) is known, and the corresponding two types of errors are given via $Y_{1(k_2)}$ and $Y_{2(k_2)}$ which are as follows :

$$Y_{1(k_2)} = k_2 Y_{2(k_2)}$$

where
$$Y_{2(k_2)} = \frac{\alpha'_{k_2} \mu}{k_2 (\alpha'_{k_2} R_1 \alpha_{k_2})^{\frac{1}{2}} + (\alpha'_{k_2} R_2 \alpha_{k_2})^{\frac{1}{2}}}$$
(3.5.13)

Similarly, the two errors denoted by e_{1B} and e_{2B} due to (α_B, c_B), where c_B is yet to be specified, are given via $Y_{1(B)}$ and $Y_{2(B)}$ respectively which are as follows :

$$Y_{1(B)} = \frac{\alpha'_B \mu - c_B}{(\alpha'_B R_1 \alpha_B)^{\frac{1}{2}}}$$

$$Y_{2(B)} = \frac{c_B}{(\alpha'_B R_2 \alpha_B)^{\frac{1}{2}}}$$

Set $\quad Y_{1(B)} = Y_{1(k_2)}$

and $\quad Y_{2(B)} = Y_{2(k_2)}$

$$\Rightarrow \frac{\alpha'_B \mu - c_B}{(\alpha'_B R_1 \alpha_B)^{\frac{1}{2}}} = \frac{k_2 \alpha'_{k_2} \mu}{k_2 (\alpha'_{k_2} R_1 \alpha_{k_2})^{\frac{1}{2}} + (\alpha'_{k_2} R_2 \alpha_{k_2})^{\frac{1}{2}}}$$
(3.5.14)

and
$$\frac{c_B}{(\underset{\sim}{\alpha}_B' R_2 \underset{\sim}{\alpha}_B)^{\frac{1}{2}}} = \frac{\underset{\sim}{\alpha}_{k_2}' \underset{\sim}{\mu}}{(\underset{\sim}{\alpha}_{k_2}' R_1 \underset{\sim}{\alpha}_{k_2})^{\frac{1}{2}} + (\underset{\sim}{\alpha}_{k_2}' R_2 \underset{\sim}{\alpha}_{k_2})^{\frac{1}{2}}}$$
(3.5.15)

(3.5.14) and (3.5.15) implies (3.5.8).

(b) Set $\dfrac{\theta_1}{k_2} = \theta$

$\Rightarrow \theta_1 = k_2 \theta$

which when put in (3.5.12) gives (3.5.9).

3.5.1 AN ILLUSTRATIVE EXAMPLE

<u>Example 3.5.1</u> : Let us consider the Example 3.3.1 again.

(a) Let $k_2 = 2.0$. Then we find $\underset{\sim}{\alpha}_{k_2}$ and consequently $Y_{1(k_2)}$ and $Y_{2(k_2)}$ which are as follows :

$$Y_{1(k_1)} = 2.0913 \quad , \quad Y_{2(k_2)} = 1.0456$$

Next we find $\underset{\sim}{\alpha}_B$, and then c_B using (3.5.8) . Thus $Y_{1(B)}$ and $Y_{2(B)}$ are given by

$$Y_{1(B)} = 2.0908 \; , \; Y_{2(B)} = 1.0454.$$

Now suppose c_B is given and $c_B = 2.5$.

Then $Y_{1(B)} = 2.5995, \; Y_{2(B)} = 0.7936$
and k_2 using (3.5.8) is given by

$$k_2 = 3.2755.$$

With this k_2 we find $\underset{\sim}{\alpha}_{k_2}$ which gives $Y_{1(k_2)}$ and $Y_{2(k_2)}$:

$$Y_{1(k_2)} = 2.6012 \; ; \; Y_{2(k_2)} = .7941$$

(b) The value of k_2 for which $\underset{\sim}{\alpha}_B = \underset{\sim}{\alpha}_{k_2}$ is 1.1925.

Remark 3.5.1 : It is interesting to note that the above theorem relates our criterion of maximizing the Bhattacharyya distance with the minimization of total probability of misclassification subject to a linear relationship between the two types of errors. Thus the procedure based on the maximization of the Bhattacharyya distance is justified.

3.6   METHOD OF OBTAINING $\underset{\sim}{\alpha}$ IN THE CASE OF LARGE SAMPLE FOR STATIONARY TIME SERIES

3.6.1   EXPRESSION FOR THE OPTIMAL $\underset{\sim}{\alpha}$

First we have listed the assumptions based on which attempts have been made to give an explicit expression for the optimal vector $\underset{\sim}{\alpha}$ in the sense of maximizing the Bhattacharyya distance asymptotically. We made the following assumptions :

A1. The n-dimensional vector $\underset{\sim}{x} = (x(o),\ldots,x(n-1))$ is covariance stationary normal time series with mean $\underset{\sim}{\mu}_j$ and covariance matrix $R_j = ((r_j(s-t),s,t = 0,\ldots,n-1))$ under hypothesis $H_j(j = 1,2)$.

A2. The spectral densities $f_j(\lambda)$ of the process under the hypotheses are positive on $[-\pi,\pi]$.

A3. The sequence of mean difference $\delta(t)$ satisfies

(i) $\sup\limits_{t} |\delta(t)| < \infty$ and

(ii) $\xi_{(n)}(\tau) \overset{\Delta}{=} \dfrac{1}{n} \sum\limits_{t=0}^{n-1-|\tau|} \delta(t+|\tau|)\delta(t)$

has a limit given by

$$\xi(\tau) \overset{\Delta}{=} \lim_{n \to \infty} \xi_{(n)}(\tau) = \int_{-\pi}^{\pi} e^{i\lambda\tau} \frac{dM(\lambda)}{2\pi} \qquad (3.6.1)$$

where $M(\lambda)$ is a monotone non-decreasing function uniquely defined by the conditions $M(-\pi) = 0$ and continuity from the right (see [59]) and $\xi(0) > 0$.

A4. The covariance sequence $r_j(t)$ satisfies

$$\sum_{t=-\infty}^{\infty} |t|^{1+\beta} |r_j(t)| < \infty \qquad (3.6.2)$$

for $j = 1,2$ and some β, $0 < \beta < 1$.

Remark 3.6.1 :   (a) It may be noted that if we take $\delta(t)$ as a stochastic process which is ergodic in autocorrelation ([43]), then $M(\lambda)$ is nothing but its spectral distribution function.

(b) (3.6.2) implies that

$$|f_j(\lambda_1) - f_j(\lambda_2)| \leq c_2 |\lambda_1 - \lambda_2|^{\beta} \qquad (3.6.2)'$$

where c_2 is some fixed constant (see [32]). $(3.6.2)'$ is a sufficient condition (see [5]) for the Fourier transform of $\{r_j(t)\}_{t=0}^{\infty}$ to converge uniformly to $f_j(\lambda)$ in $[-\pi, \pi]$.

Under the assumptions stated above the following theorem is a logical continuation of **Remark 3.3.11**.

Theorem 3.6.1 : Suppose the assumptions Al to A4 are satisfied. If $f_\theta(\lambda)$, defined by $f_\theta(\lambda) \overset{\Delta}{=} f_1(\lambda) - \theta f_2(\lambda)$ for $\lambda \epsilon [-\pi, \pi]$, is positive, then

$$\lim_{n \to \infty} \{ - \frac{1}{n} \ln \rho_2(1,2;y_\theta) \}$$

$$= \frac{1}{4} [\int_{-\pi}^{\pi} \frac{dM(\lambda)}{2\pi f_\theta(\lambda)}]^2 / [\int_{-\pi}^{\pi} \frac{f_1(\lambda) + f_2(\lambda)}{f_\theta^2(\lambda)} \frac{dM(\lambda)}{2\pi}]$$

$$(3.6.3)$$

where y_θ is defined in **Remark 3.3.10**. We need the following lemmas to prove the theorem.

Lemma 3.6.1 : Assume $R_\theta \overset{\Delta}{=} R_1 - \theta R_2$ is positive definite; this entails $f_\theta(\lambda) > 0$ on $[-\pi, \pi]$. Then

$$\lim_{n \to \infty} \frac{1}{n} \underset{\sim}{\delta}' R_\theta^{-1} \underset{\sim}{\delta} = \int_{-\pi}^{\pi} \frac{1}{f_\theta(\lambda)} \frac{dM(\lambda)}{2\pi}$$

Proof : See Appendix E.

Lemma 3.6.2 : $\lim_{n \to \infty} \frac{1}{n} \underset{\sim}{\delta}' R_\theta^{-1} R_j R_\theta^{-1} \underset{\sim}{\delta} = \int_{-\pi}^{\pi} \frac{f_j(\lambda)}{f_\theta^2(\lambda)} \frac{dM(\lambda)}{2\pi}$ $(j = 1,2)$.

Proof : See Appendix E.

Proof of the Theorem 3.6.1 : We have from (3.3.46),

$$\rho_2(1,2;y_\theta) = \frac{(\sigma_1^2 \sigma_2^2)^{\frac{1}{4}}}{\{ \frac{1}{2}(\sigma_1^2 + \sigma_2^2) \}^{\frac{1}{2}}} \exp\{ - \frac{1}{4} \frac{(\underset{\sim}{\delta}' R_\theta^{-1} \underset{\sim}{\delta})^2}{\sigma_1^2 + \sigma_2^2} \}$$

where $\quad \sigma_j^2 \overset{\Delta}{=} \underset{\sim}{\delta}'R_\theta^{-1} R_j R_\theta^{-1}\underset{\sim}{\delta} \quad (j = 1,2).$

Thus,

$$-\frac{1}{n}\ln\rho_2(1,2;Y_\theta) = -\frac{1}{4n}\ln\sigma_1^2 - \frac{1}{4n}\ln\sigma_2^2 + \frac{1}{2n}\ln\frac{1}{2}(\sigma_1^2+\sigma_2^2) + \frac{1}{4n}\frac{(\underset{\sim}{\delta}'R_\theta^{-1}\underset{\sim}{\delta})^2}{\sigma_1^2+\sigma_2^2}$$

$$= -\frac{1}{4}\left[\frac{\ln n}{n} + \frac{1}{n}\ln\frac{\sigma_1^2}{n}\right] - \frac{1}{4}\left[\frac{\ln n}{n} + \frac{1}{n}\ln\frac{\sigma_2^2}{n}\right]$$

$$+ \frac{1}{2}\left[\frac{\ln n}{n} + \frac{1}{n}\ln\{\frac{\frac{1}{2}(\sigma_1^2 + \sigma_2^2)}{n}\}\right]$$

$$+ \frac{1}{4}\{\frac{\underset{\sim}{\delta}'R_\theta^{-1}\underset{\sim}{\delta}}{n}\}^2/\{\frac{1}{n}(\sigma_1^2 + \sigma_2^2)\}$$

$$= -\frac{1}{4n}\ln\frac{\sigma_1^2}{n} - \frac{1}{4n}\ln\frac{\sigma_2^2}{n} + \frac{1}{2n}\ln\{\frac{\frac{1}{2}(\sigma_1^2+\sigma_2^2)}{n}\}$$

$$+ \frac{1}{4}\{\frac{\underset{\sim}{\delta}'R_\theta^{-1}\underset{\sim}{\delta}}{n}\}^2/\{\frac{1}{n}(\sigma_1^2+\sigma_2^2)\}$$

$$\tag{3.6.4}$$

The first three terms in (3.6.4) converge to zero as $n \to \infty$, by **Lemma 3.6.2.** The last term in (3.6.4) converges to the indicated value by **Lemma 3.6.1** and **Lemma 3.6.2.** Hence the Theorem is proved.

Define

$$G(\theta) \overset{\Delta}{=} [\int_{-\pi}^{\pi}\frac{1}{f_\theta(\lambda)}\frac{dM(\lambda)}{2\pi}]^2/[\int_{-\pi}^{\pi}\frac{f_1(\lambda)+f_2(\lambda)}{f_\theta^2(\lambda)}\frac{dM(\lambda)}{2\pi}]$$

$$\tag{3.6.5}$$

The following theorem characterizes the value of θ for which $-\ln\rho_2(1,2;Y_\theta)$ has a maximum for large sample .

$$-\hat{\theta}_n = \frac{\{\underset{\sim}{\alpha}'_* \underset{\sim}{\delta}/\underset{\sim}{\alpha}'_*(R_1+R_2)\underset{\sim}{\alpha}_*\}^2 + 1/\underset{\sim}{\alpha}'_* R_2\underset{\sim}{\alpha}_* - 2/\underset{\sim}{\alpha}'_*(R_1+R_2)\underset{\sim}{\alpha}_*}{\{\underset{\sim}{\alpha}'_*\underset{\sim}{\delta}/\underset{\sim}{\alpha}'_*(R_1+R_2)\underset{\sim}{\alpha}_*\}^2 + 1/\underset{\sim}{\alpha}'_* R_1\underset{\sim}{\alpha}_* - 2/\underset{\sim}{\alpha}'_*(R_1+R_2)\underset{\sim}{\alpha}_*} \quad (3.6.8)$$

converges to 1 as $n \to \infty$. In fact, (3.6.8) can be put in the following form (after plugging in the value of $\underset{\sim}{\alpha}_*$)

$$\frac{1 + \frac{1}{n} \dfrac{1}{\frac{1}{n}\underset{\sim}{\delta}'(R_1+R_2)^{-1}R_2(R_1+R_2)^{-1}\underset{\sim}{\delta}} - \dfrac{2/n}{\frac{1}{n}\underset{\sim}{\delta}'(R_1+R_2)^{-1}\underset{\sim}{\delta}}}{1 + \frac{1}{n} \dfrac{1}{\frac{1}{n}\underset{\sim}{\delta}'(R_1+R_2)^{-1}R_1(R_1+R_2)^{-1}\underset{\sim}{\delta}} - \dfrac{2/n}{\frac{1}{n}\underset{\sim}{\delta}'(R_1+R_2)^{-1}\underset{\sim}{\delta}}}$$

which converges to 1, by **Lemma 3.6.1** and **Lemma 3.6.2**.

Remark 3.6.3 : $\underset{\sim}{\alpha}_*$ alone cannot specify a test. The value of c of a linear procedure is to be known in order to obtain a complete description of the test. This can be done in the following ways.

1) If we specify one kind of error (say, e_1), then c is automatically known. One observation can be made in this connection. Suppose e_1 is given or equivalently y_1. Then

$$e_2 = 1 - \Phi\left(\frac{\underset{\sim}{\alpha}'_*\underset{\sim}{\delta} - y_1(\underset{\sim}{\alpha}'_* R_1\underset{\sim}{\alpha}_*)^{\frac{1}{2}}}{(\underset{\sim}{\alpha}'_* R_2\underset{\sim}{\alpha}_*)^{\frac{1}{2}}}\right)$$

$$= 1 - \Phi\left(\frac{\underset{\sim}{\delta}'(R_1+R_2)^{-1}\underset{\sim}{\delta} - y_1(\underset{\sim}{\delta}'(R_1+R_2)^{-1}R_1(R_1+R_2)^{-1}\underset{\sim}{\delta})^{\frac{1}{2}}}{(\underset{\sim}{\delta}'(R_1+R_2)^{-1}R_2(R_1+R_2)^{-1}\underset{\sim}{\delta})^{\frac{1}{2}}}\right)$$

$$= 1 - \Phi\left(\frac{(n)^{\frac{1}{2}}\left(\frac{\delta'(R_1+R_2)^{-1}\delta}{n}\right) - Y_1\left(\frac{1}{n}\delta'(R_1+R_2)^{-1}R_1(R_1+R_2)^{-1}\delta\right)^{\frac{1}{2}}}{\left(\frac{1}{n}\delta'(R_1+R_2)^{-1}R_2(R_1+R_2)^{-1}\delta\right)^{\frac{1}{2}}}\right)$$

$$\xrightarrow[(n \to \infty)]{} 0 \ , \text{ by Lemma 3.6.1 and Lemma 3.6.2.}$$

2) One can specify c in such a way that the total error of misclassification is a minimum.

3) Each procedure is evaluated in terms of the two probabilities of misclassification. One procedure is better than another if each probability of misclassification of the former is not greater than the corresponding one of the latter and atleast one is less. A procedure is admissible if there is no other procedure which is better.

The following theorem which characterizes an admissible procedure is given in Bahadur and Anderson ([4]).

Theorem 3.6.3 : A linear procedure with

$$\alpha = (\tau_1 R_1 + \tau_2 R_2)^{-1}\delta \qquad (3.6.9)$$

$$c = \alpha'\mu_1 - \tau_1(\alpha'R_1\alpha) \qquad (3.6.10)$$

$$= \alpha'\mu_2 + \tau_2(\alpha'R_2\alpha)$$

for any τ_1, τ_2 such that $\tau_1 R_1 + \tau_2 R_2$ is positive definite is admissible.

Given τ_1, τ_2 such that $\tau_1 R_1 + \tau_2 R_2$ is p.d., one would compute the optimal $\underset{\sim}{\alpha}$ satisfying $(\tau_1 R_1 + \tau_2 R_2)\underset{\sim}{\alpha} = \underset{\sim}{\delta}$ and then compute c as given in (3.6.10). Usually, τ_1 and τ_2 are not given. We may specify them via the maximization of the Bhattacharyya distance. For large sample, we can take $\tau_1 = \tau_2 = 1$ and the procedure is

$$\underset{\sim}{\alpha}_* = (R_1 + R_2)^{-1}\underset{\sim}{\delta} \qquad (3.6.11)$$

$$\left.\begin{array}{l} c_* = \underset{\sim}{\delta}'(R_1+R_2)^{-1}\underset{\sim}{\mu}_1 - \underset{\sim}{\delta}'(R_1+R_2)^{-1}R_1(R_1+R_2)^{-1}\underset{\sim}{\delta} \\[2mm] \quad = \underset{\sim}{\delta}'(R_1+R_2)^{-1}\underset{\sim}{\mu}_2 + \underset{\sim}{\delta}'(R_1+R_2)^{-1}R_2(R_1+R_2)^{-1}\underset{\sim}{\delta} \end{array}\right\} \quad (3.6.12)$$

Now we observe the following.

If (e_{1*}, e_{2*}) are the two probabilities of misclassification resulting from the use of the linear procedure defined by (3.6.11) and (3.6.12), then by (3.3.27),

$$e_{1*} = 1 - \Phi\left(\frac{\underset{\sim}{\alpha}'\underset{\sim}{\mu}_1 - c_*}{(\underset{\sim}{\alpha}_*'R_1\underset{\sim}{\alpha}_*)^{\frac{1}{2}}}\right)$$

$$\quad = 1 - \Phi\left((\underset{\sim}{\delta}'(R_1+R_2)^{-1}R_1(R_1+R_2)^{-1}\underset{\sim}{\delta})^{\frac{1}{2}}\right)$$

$$\quad = 1 - \Phi\left[(n)^{\frac{1}{2}}\left(\frac{1}{n}\underset{\sim}{\delta}'(R_1+R_2)^{-1}R_1(R_1+R_2)^{-1}\underset{\sim}{\delta}\right)^{\frac{1}{2}}\right]$$

$$\underset{(n\to\infty)}{\longrightarrow} 0, \text{ by Lemma 3.6.2.}$$

By a similar argument, $e_{2*} \longrightarrow 0$ as $n \to \infty$.

3.6.2 EXAMPLES

In this section, we give some examples to illustrate the theory presented in the previous section. Examples are chosen so as to satisfy the assumptions A1 to A4 of Section 3.6.1.

Let $\{Z(t), t \geq 0\}$ be an autoregressive normal process of order 2 (abbreviated as AR(2)) i.e. $Z(t)$ satisfies

$$Z(t) = \beta_1 Z(t-1) + \beta_2 Z(t-2) + \varepsilon(t) \qquad (3.6.13)$$

where $\{\varepsilon(t), t \geq 0\}$ is a normal process with

$$E\varepsilon(t) = 0$$

and

$$\text{Cov}(\varepsilon(t), \varepsilon(t+\tau)) = \begin{cases} 0 & \tau = 1,2,\ldots \\ 1 & \tau = 0. \end{cases}$$

Define $Z(t) \overset{\Delta}{=} X(t) - \mu(t)$, where $EX(t) = \mu(t)$. It is well known ([12]) that the process $\{Z(t), t \geq 0\}$ is stationary provided $|\xi_i| < 1$, $(i = 1,2)$, where $\{\xi_i\}$ are the roots of the equation

$$m^2 - \beta_1 m - \beta_2 = 0 \qquad (3.6.14)$$

A sufficient condition in terms of β_1, β_2 for stationarity of $\{Z(t), t \geq 0\}$ is that β_1, β_2 should lie in a triangular region :

$$\left. \begin{array}{c} \beta_1 + \beta_2 < 1 \\ \beta_2 - \beta_1 < 1 \\ -1 < \beta_2 < 1 \end{array} \right\} \qquad (3.6.15)$$

It may be verified that under conditions (3.6.15), $|\xi_1| < 1$ and $|\xi_2| < 1$ (see [10]). Thus $\{Z(t), t \geq 0\}$ satisfies the assumption A1 provided (3.6.15) holds.

The spectral density of AR(2) is given by ([10])

$$f(\lambda) = \frac{1}{2\pi} \frac{1}{|1 - \beta_1 e^{-i\lambda} - \beta_2 e^{-2i\lambda}|^2} > 0 \text{ on } [-\pi, \pi].$$

Hence A2 is satisfied.

Take $\delta(t) = \cos\frac{\pi}{2}t$

$$(3.6.16)$$

Then, $\dfrac{1}{n} \displaystyle\sum_{t=0}^{n-1-|\tau|} \delta(t+|\tau|)\delta(t)$

$$= \frac{1}{n} \sum_{t=0}^{n-1-|\tau|} \{\cos\frac{\pi}{2}(t+|\tau|)\cos\frac{\pi}{2}t\}$$

$$= \frac{1}{2n} \sum_{t=0}^{n-1-|\tau|} \{\cos\frac{\pi}{2}(2t+|\tau|) + \cos\frac{\pi}{2}(|\tau|)\}$$

$$= \frac{1}{2n}\left[(n-|\tau|)\cos\frac{\pi}{2}(|\tau|) + \sum_{t=0}^{n-1-|\tau|}\cos\frac{\pi}{2}(2t+|\tau|)\right]$$

$$= \frac{1}{2n}\left[(n-|\tau|)\cos\frac{\pi}{2}|\tau| + \frac{1}{2}\sum_{t=0}^{n-1-|\tau|}\{e^{i\pi(t+\frac{|\tau|}{2})}\right.$$

$$\left. + e^{-i\pi(t+\frac{|\tau|}{2})}\}\right]$$

$$= \frac{1}{2n}\left[(n-|\tau|)\cos\frac{\pi}{2}|\tau| + \frac{1}{2}e^{i\pi\frac{|\tau|}{2}}\left(\sum_{t=0}^{n-1-|\tau|}\cos\pi t\right)\right.$$

$$\left. + \frac{1}{2}e^{-i\pi\frac{|\tau|}{2}}\left(\sum_{t=0}^{n-1-|\tau|}\cos\pi t\right)\right]$$

$$= \frac{1}{2}\left[(1- \frac{|\tau|}{n})\cos \frac{\pi}{2} |\tau| + \frac{1}{2} e^{i\pi \frac{|\tau|}{2}} \left(\frac{\sum_{t=o}^{n-1-|\tau|} (-1)^t}{n} \right) \right.$$

$$\left. + \frac{1}{2} e^{-i\pi \frac{|\tau|}{2}} \left(\frac{\sum_{t=o}^{n-1-|\tau|} (-1)^t}{n} \right) \right]$$

This implies, $\lim_{n \to \infty} \frac{1}{n} \sum_{t=o}^{n-1-|\tau|} \delta(t+|\tau|)\delta(t) = \frac{1}{2} \cos \frac{\pi}{2} |\tau|$

Now, take $M(\lambda)$ to be a step function having jumps at $\pm \frac{\pi}{2}$ of height $\frac{\pi}{2}$.

Then $\int_{-\pi}^{\pi} e^{i\lambda\tau} \frac{dM(\lambda)}{2\pi} = \frac{1}{2} \cos \frac{\pi}{2} \tau$

Thus, $\lim_{n \to \infty} \frac{1}{n} \sum_{t=o}^{n-1-|\tau|} \delta(t+|\tau|)\delta(t) = \int_{-\pi}^{\pi} e^{i\lambda\tau} \frac{dM(\lambda)}{2\pi}$

Hence, A3 is satisfied.

Let $r(h) = EZ(t)Z(t+h)$.

Then it can be shown that $r(h)$ satisfies the difference equation

$$r(h) = \beta_1 r(h-1) + \beta_2 r(h-2), \qquad (3.6.17)$$

the solution of which is given by ([10]),

$$r(h) = \frac{(1-\xi_2^2)\xi_1^{h+1}}{(\xi_1-\xi_2)(1+\xi_1\xi_2)} + \frac{(1-\xi_1^2)\xi_2^{h+1}}{(\xi_2-\xi_1)(1+\xi_1\xi_2)} , \quad (3.6.18)$$

$$(h = 0,1,2,\dots)$$

where ξ_i's are the roots of $(3.6.14)$.

Write $r(h) = a_1 \xi_1^{h+1} + a_2 \xi_2^{h+1}$,

where

$$a_1 = \frac{(1-\xi_2^2)}{(\xi_1 - \xi_2)(1 +\xi_1\xi_2)} , \quad a_2 = \frac{(1-\xi_1^2)}{(\xi_2-\xi_1)(1+\xi_1\xi_2)}$$

Then

$$\sum_{t=0}^{\infty} |t|^{1+\beta} |\dot{r}(t)|$$

$$\leq |a_1| \left(\sum_{t=0}^{\infty} |t|^{1+\beta} |\xi_1|^{t+1} \right) + |a_2| \left(\sum_{t=0}^{\infty} |t|^{1+\beta} |\xi_2|^{t+1} \right)$$

Letting $u_t = t^{1+\beta} |\xi_1|^{t+1}$, we have

$$\lim_{t \to \infty} \frac{u_{t+1}}{u_t} = \lim_{t \to \infty} \left(1 + \frac{1}{t} \right)^{1+\beta} |\xi_1|$$

$$= |\xi_1| < 1$$

Thus by D'Alembert's ratio test, $\sum_t |t|^{1+\beta} |\xi_1|^{t+1} < \infty$.

Similarly, we have, $\sum_t |t|^{1+\beta} |\xi_2|^{t+1} < \infty$.

Hence A4 is satisfied.

Example 3.6.1 : Let

$$H_1 : Z(t) = 0.5\, Z(t-1) + 0.3\, Z(t-2) + \varepsilon(t)$$

$$H_2 : Z(t) = 0.5\, Z(t-1) - 0.3\, Z(t-2) + \varepsilon(t)$$

Let

$$E_{H_2} X(t) = 0, \quad E_{H_1} X(t) = \delta(t) = \cos \frac{\pi}{2} t$$

The first row of R_1 is given by (for n = 25) ;

(1.0,0.7143,0.6571,0.5428,0.4685,0.39712,0.33912,0.28870,0.2460,

0.20965,0.17865,0.15222,0.12970,0.11052,0.09417,0.08024,0.06837,

0.05825,0.04964,0.04229,0.036041,0.03071,0.026167,0.022291,

0.01899).

The first row of R_2 is given by (for n = 25):

(1.0,0.385,-0.108,-0.1695,-.0523,0.02467,0.02804,

0.00661, -0.0051, -0.004537, -0.000737, 0.00099,0.00072,

0.0000, -0.0002,-0.00011,0.0000,0.0000,0.0000,0.0000,

0.0000,0.0000,0.0000,0.0000,0.0000).

We have solved the implicit equation (3.3.6) for n = 10,11,...,25. The pertinent results are shown in Table 3.18 . It is clear from the table that approximating the solution of (3.3.6) by $\underset{\sim}{\alpha} = (R_1 + R_2)^{-1}\underset{\sim}{\delta}$ becomes more and more accurate as n becomes larger and larger. See also Fig. 3.23

Similar conclusion can be drawn for the examples that follow.

Example 3.6.2 : Here we change only R_2, viz.,

$$H_1 : Z(t) = 0.5\ Z(t-1) + 0.3\ Z(t-2) + \varepsilon(t)$$
$$H_2 : Z(t) = 0.2\ Z(t-1) - 0.5\ Z(t-2) + \varepsilon(t)$$

The first row of R_2 is given by (for n = 20) :

(1.0,0.1333,-0.4733,-0.27666, 0.00366, 0.08483,0.04131,

-0.00479,-0.01479,-0.00595,0.00145,0.002516,0.00082,-0.00034,

-0.000418,-0.000106, 0.0000,0.0000,0.0000,0.0000).

The results are shown in Table 3.19. See also Fig. 3.24.

In the following examples we omit tables as well as covariance matrices for them. We give only pertinent information through figures.

Example 3.6.3 : Let

$$H_1 : \beta_1 = .5, \beta_2 = .3$$
$$H_2 : \beta_1 = -.5, \beta_2 = -.3$$

and $\qquad \delta(t) = \cos \frac{\pi}{2} t$

See Fig. 3.25.

Example 3.6.4 : Let H_1 and H_2 be the same as in Example 3.6.3. and $\qquad \delta(t) = (2)^{1/2} \cos \frac{\pi}{2} t$.

See Fig. 3.26.

Remark 3.6.4 : The convergence of θ to -1 is more rapid in Example 3.6.4 than in Example 3.6.3.

In what follows we consider examples based on covariance stationary AR(1) processes defined by

$$Z(t+1) = \rho Z(t) + \varepsilon(t+1), \text{for } |\rho| < 1, \ t = 0,1,2,\ldots .$$

It can be easily verified that it satisfies the assumptions A1 to A4.

Example 3.6.5 : Let

$$H_1 : \rho = .2, \text{ and } H_2 : \rho = .5$$
$$\delta(t) = \sin \frac{\pi}{2} t .$$

See Fig. 3.27.

Example 3.6.6 : Let

$$H_1 : \rho = -.5 ; H_2 : \rho = .8$$
$$\delta(t) = \sin \frac{\pi}{2} t$$

See Fig. 3.28.

Example 3.6.7 : Let

$$H_1 : \rho = -.5 ; H_2 : \rho = .8, \ \delta(t) = (2)^{\frac{1}{2}} \sin \frac{\pi}{2} t$$

See Fig. 3.29.

Table 3.18

Computations in Example 3.6.1

The number of observation (n)	The value of θ for which the involved iteration ends	The number of iteration required	The initial value of $\underset{\sim}{\alpha}$ for each iteration for all n
10	−0.4684516	4	
11	−0.4872228	4	
12	−0.5169756	4	
13	−0.5319041	4	
14	−0.5567487	4	
15	−0.5695292	4	
16	−0.5904840	4	
17	−0.6013405	3	$\underset{\sim}{\alpha}_o = (R_1 + R_2)^{-1} \underset{\sim}{\delta}$
18	−0.6192849	3	
19	−0.6284942	3	
20	−0.6431151	3	
21	−0.6519033	3	
22	−0.6648376	3	
23	−0.6722167	3	
24	−0.6837141	3	
25	−0.6906160	3	

Table 3.19

Computations in Example 3.6.2

The number of observation (n)	The value of θ for which the involved iteration ends	The number of iteration required	The initial value of $\underset{\sim}{\alpha}$ for each iteration for all n
2	0.0403760	6	
3	−0.0760228	2	
4	−0.0810608	4	
5	−0.0961092	4	
6	−0.1168525	4	
7	−0.1486499	5	
8	−0.1678503	5	
9	−0.2012634	5	
10	−0.2148456	5	
11	−0.2407287	5	
12	−0.2532185	5	$\underset{\sim}{\alpha}_0 = (R_1 + R_2)^{-1} \underset{\sim}{\delta}$
13	−0.2773068	4	
14	−0.2882282	4	
15	−0.3103702	4	
16	−0.3198107	4	
17	−0.3398194	4	
18	−0.3482509	4	
19	−0.3667486	4	
20	−0.3742574	4	

Fig.3.23 Example 3.6.1 (AR (2)).

$$H_1 : \beta_1 = 0.5 , \beta_2 = 0.3$$
$$H_2 : \beta_1 = 0.2 , \beta_2 = -0.5$$
$$\delta(t) = Cos \frac{\pi}{2} t$$

Fig. 3.24 Example 3.6.2 (AR (2)).

$H_1 : \beta_1 = 0.5, \ \beta_2 = 0.3$

$H_2 : \beta_1 = 0.2, \ \beta_2 = -0.5$

$\delta(t) = \cos \frac{\pi}{2} t$

Fig. 3.24 Example 3.6.2 (AR (2)) .

$H_1 : \beta_1 = 0.5, \; \beta_2 = 0.3$

$H_2 : \beta_1 = -0.5, \; \beta_2 = -0.3$

$\delta(t) = \cos \frac{\pi}{2} t$
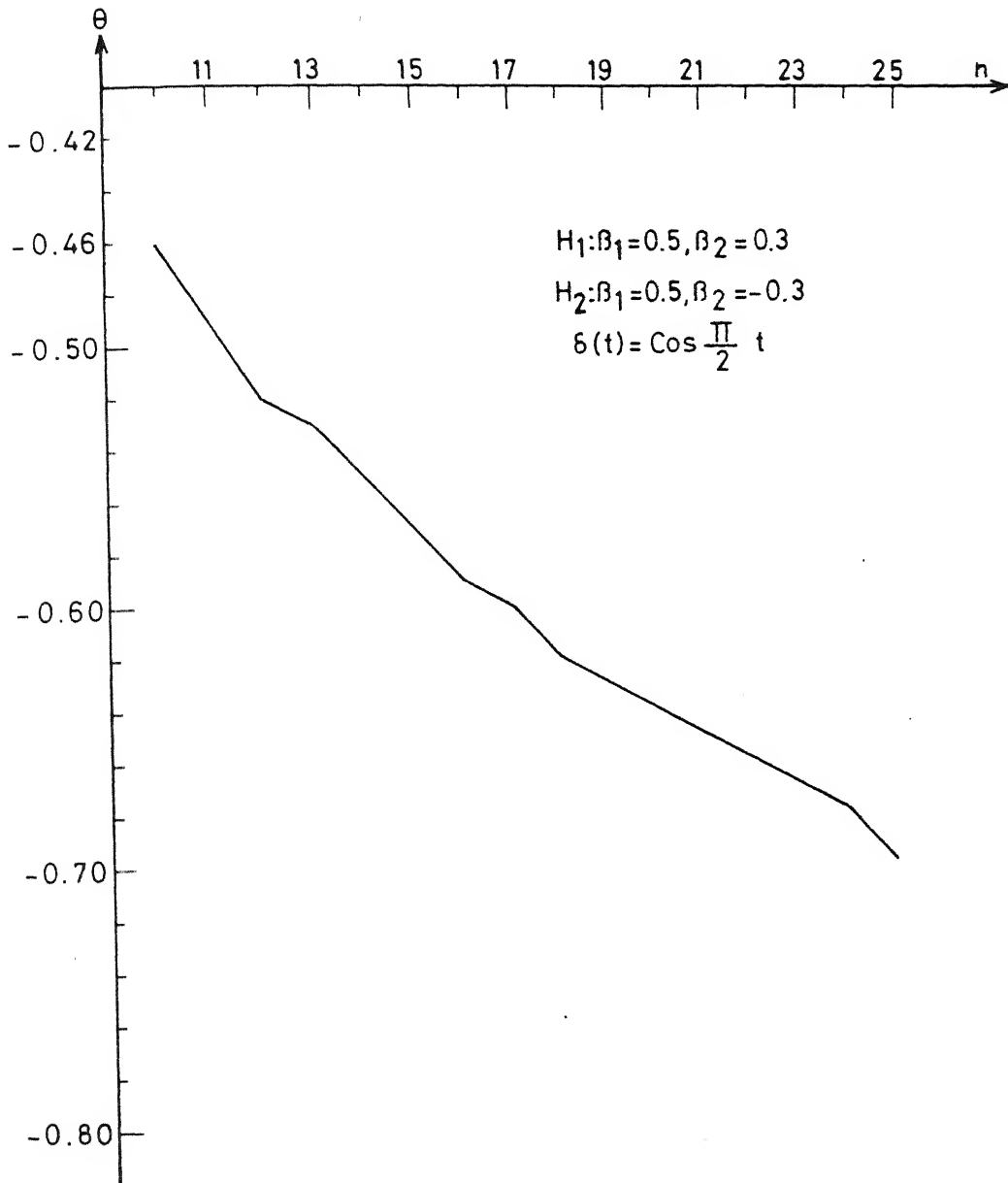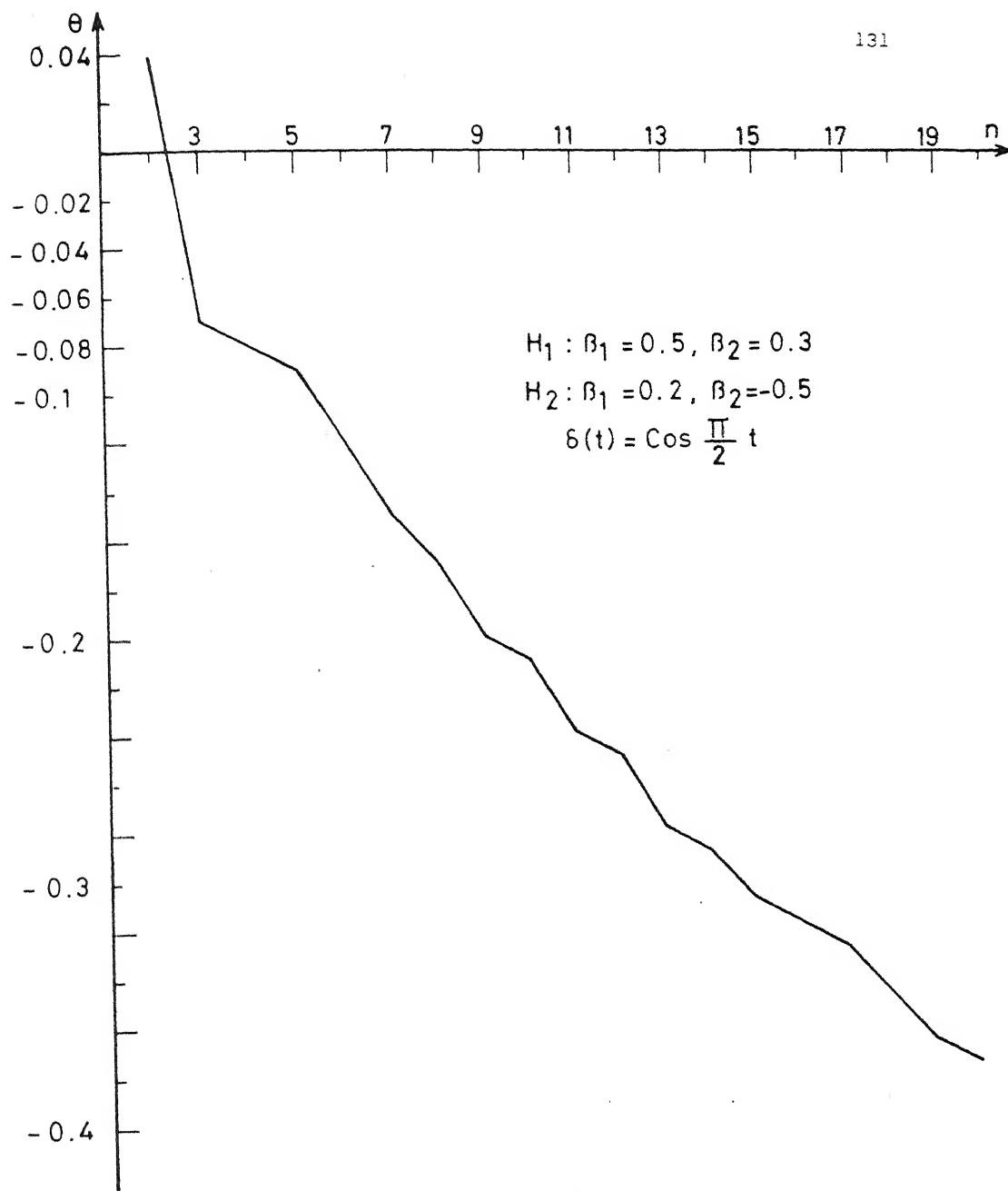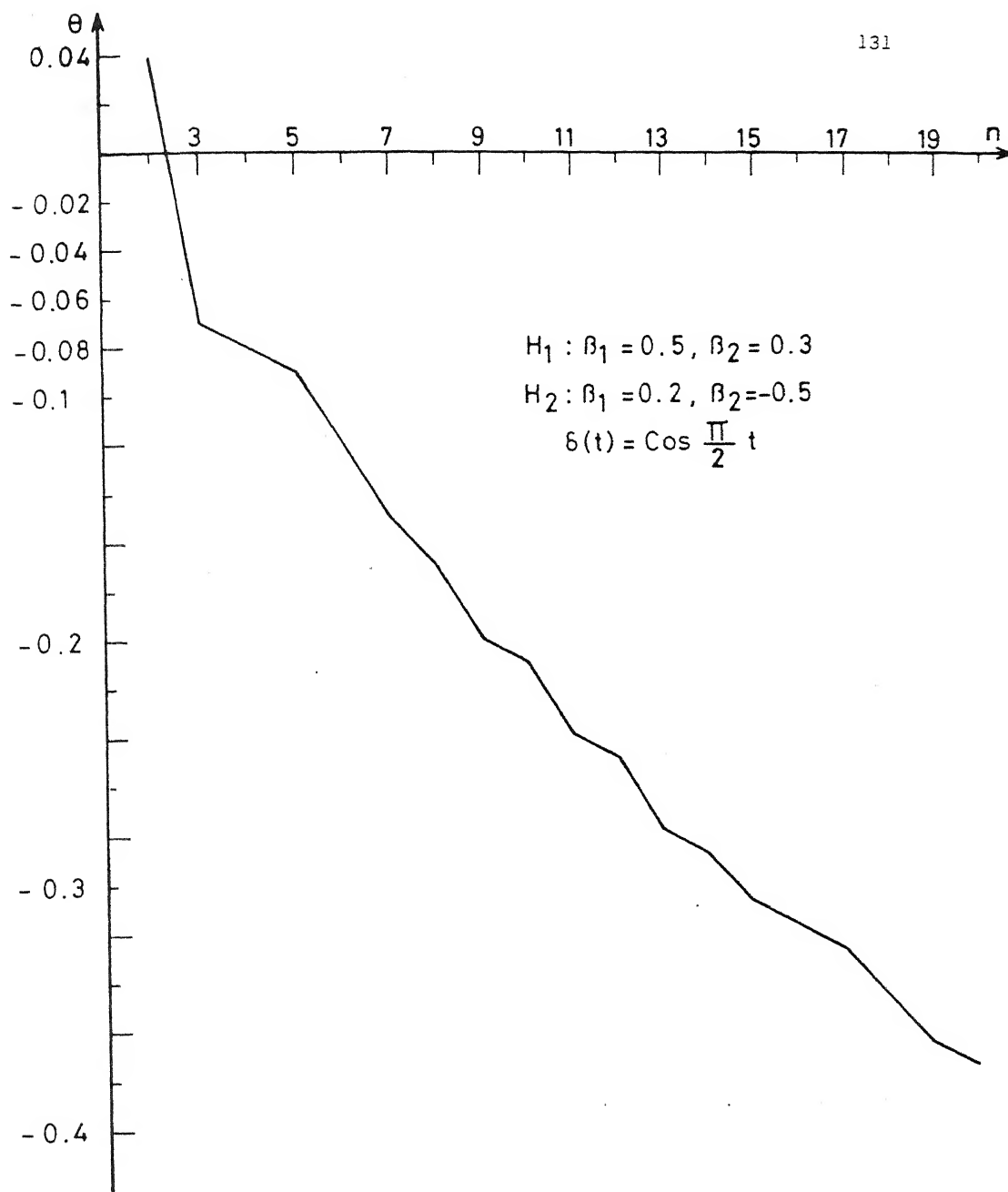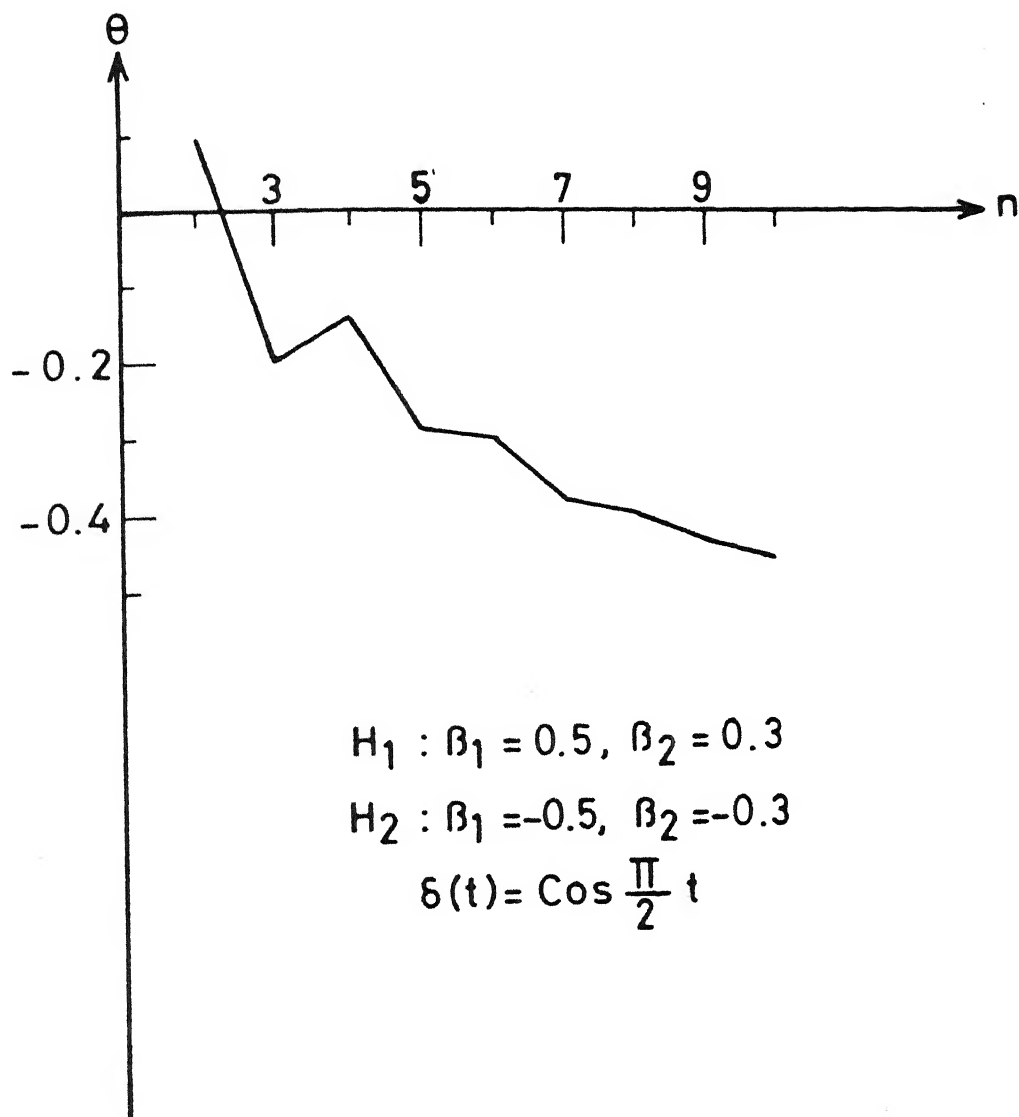
Fig. 3.25 Example 3.6.3 (A R (2)).
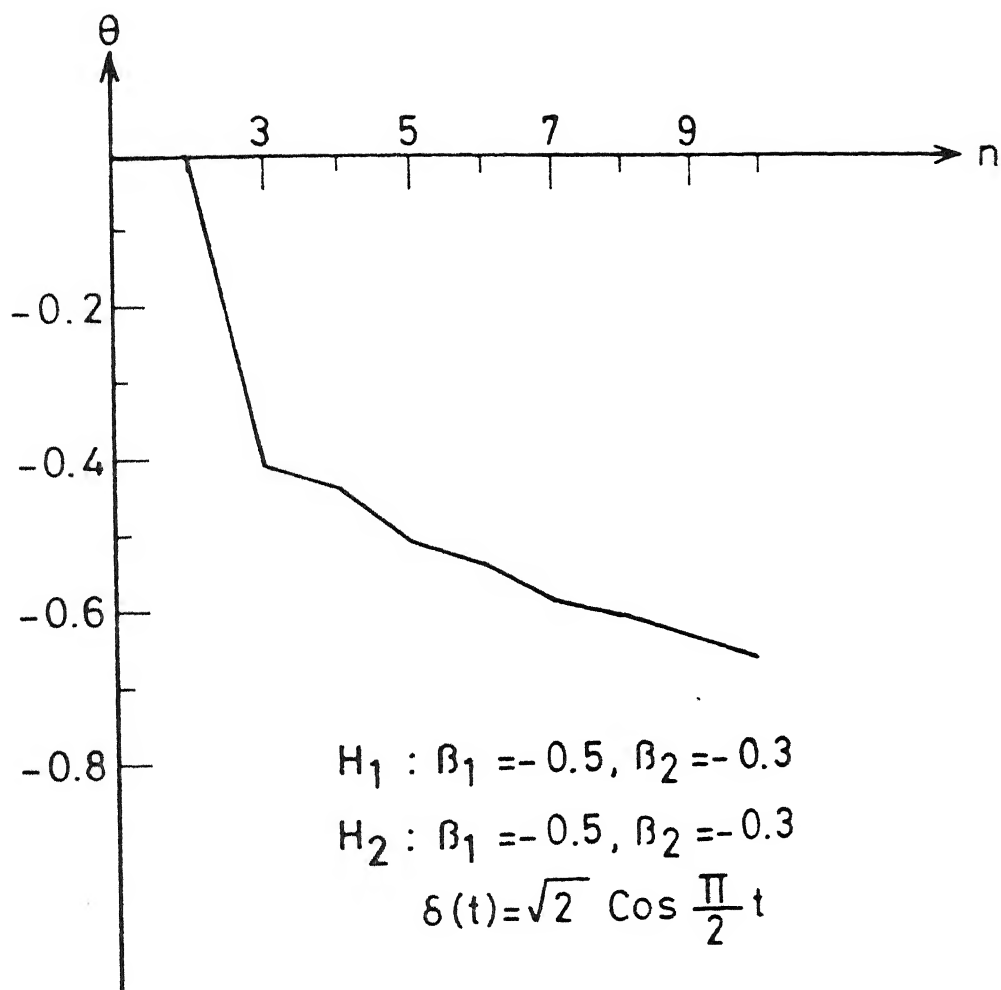
Fig. 3.26 Example 3.6.4 (AR(2)).

θ

n

$H_1: \rho = 0.2$

$H_2: \rho = 0.5$

$\delta(t) = \sin \dfrac{\pi}{2} t$
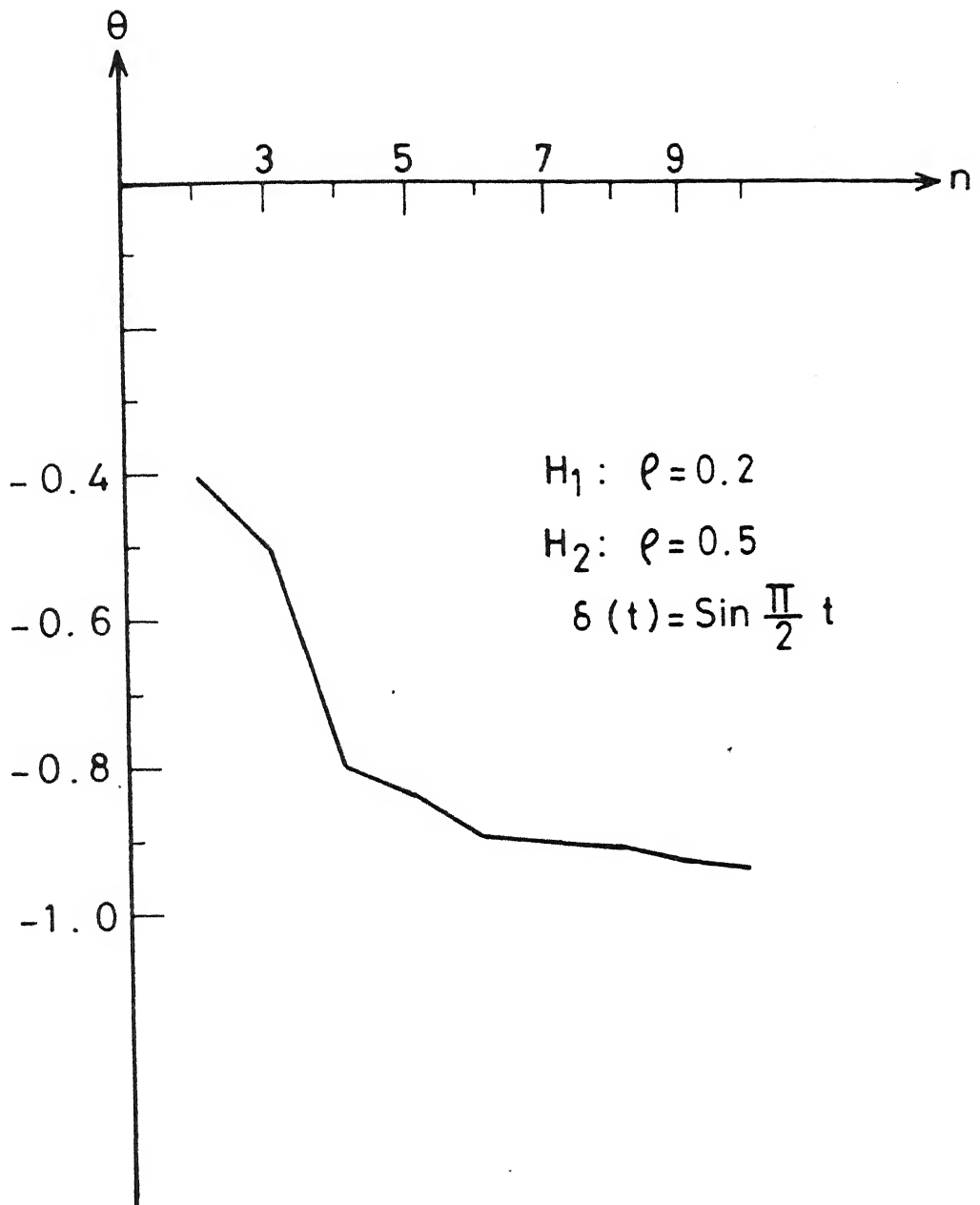
Fig.3.27 Example 3.6.5 (A R (1)).

$H_1 : \rho = -0.5$

$H_2 : \rho = 0.8$

$\delta(t) = \text{Sin} \dfrac{\pi}{2} t$
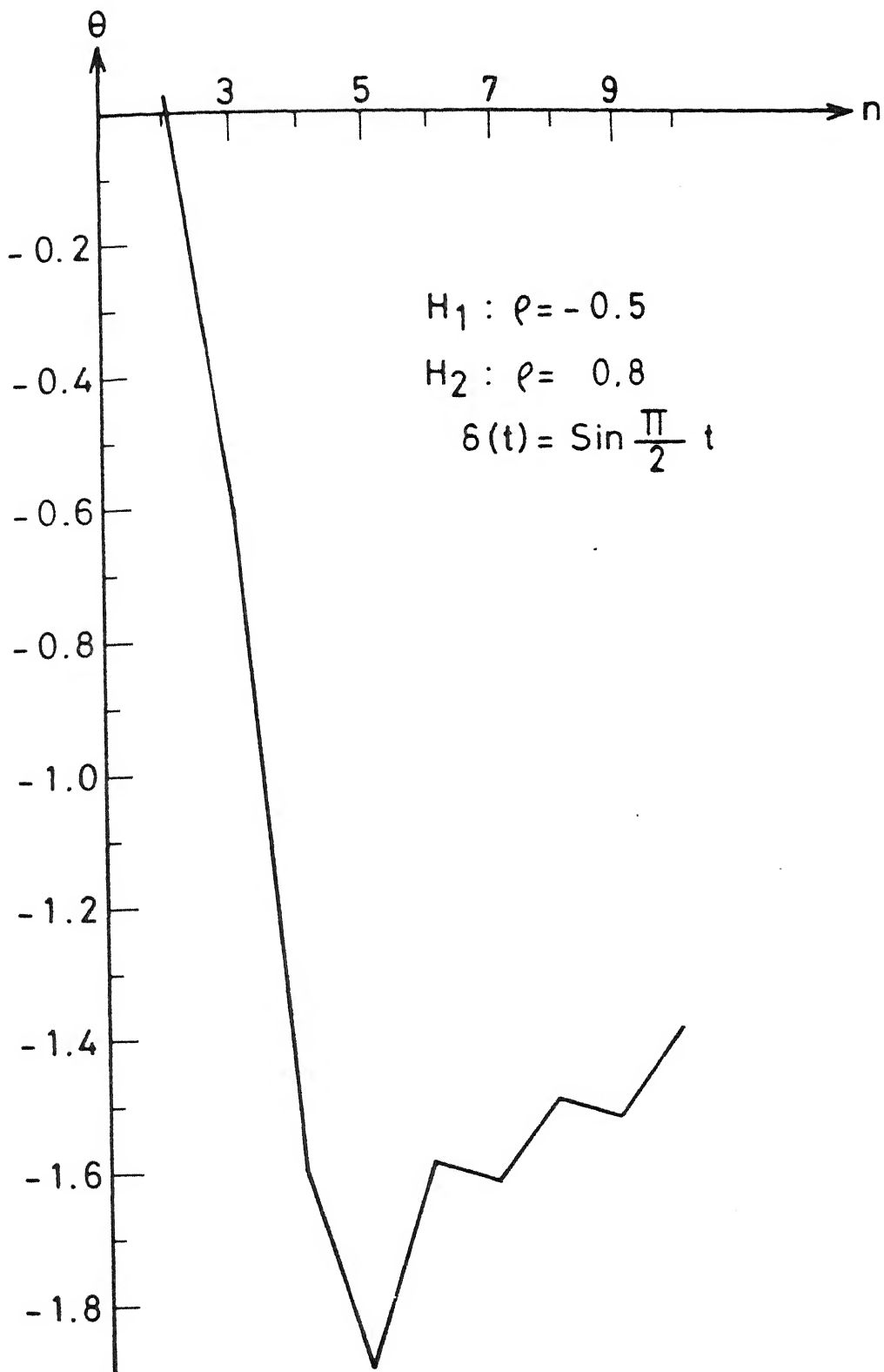
Fig. 3.28 Example 3.6.6 (AR (1)).
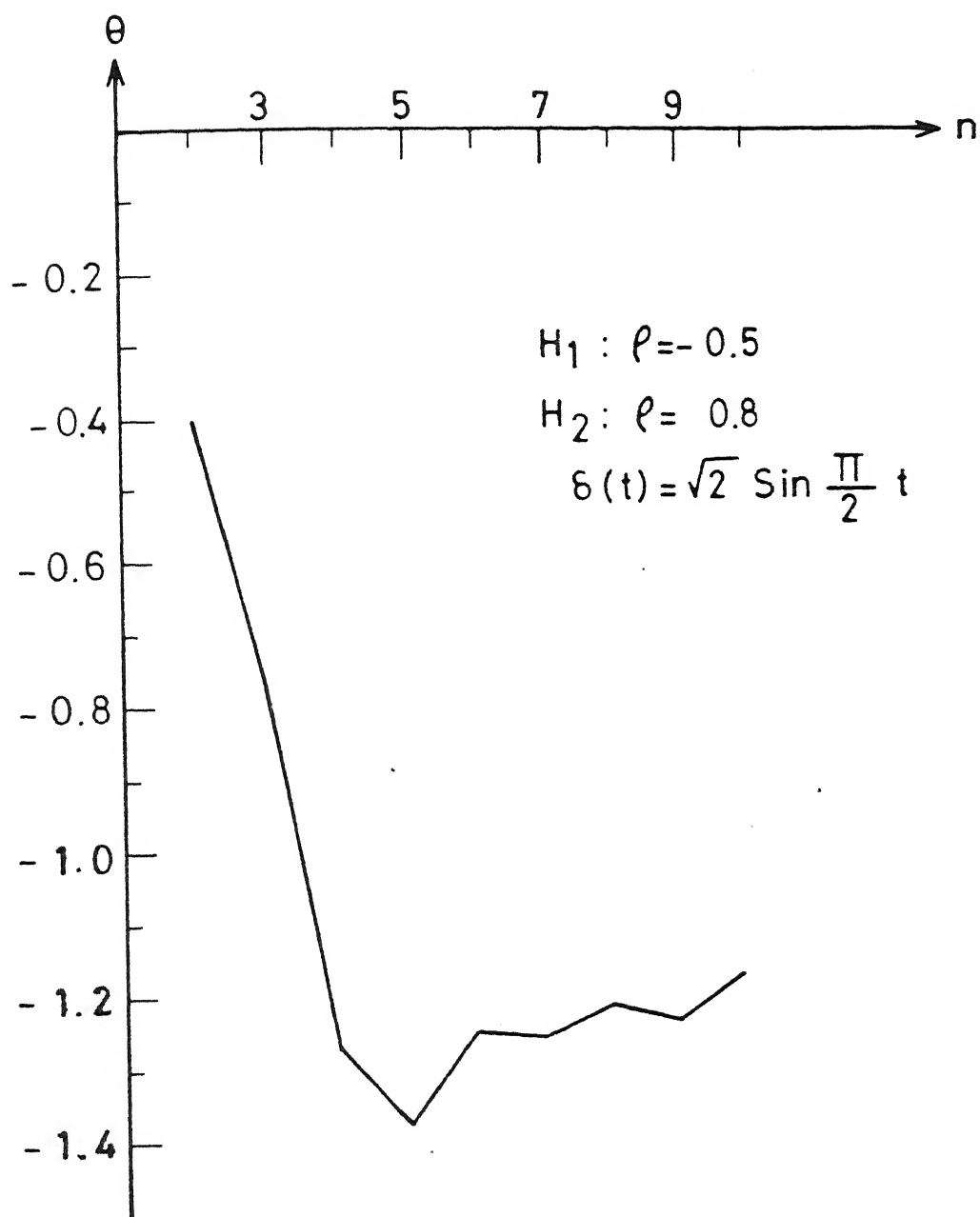
Fig.3.29 Example 3.6.7 (AR (1)).

3.7 CONCLUSION

We see in this chapter that our criterion of maximizing the Bhattacharyya distance for optimal LDF leads to solving an implicit equation for a discrete time series. By comparison with other LDFs and QDF , we note that the distance of our interest is worth-considering. In the case of stationary time series, we observe that when no explicit analytical expression is available for optimal LDFs for the criteria considered so far in the literature, the maximization of the Bhattacharyya distance does give one for large sample.

CHAPTER IV

LINEAR DISCRIMINANT FUNCTIONS FOR CONTINUOUS-TIME SERIES

4.1 INTRODUCTION

In Chapter III, the process $\{x(t), t\varepsilon T\}$ of our interest was discrete in time. Now, we assume the parameter set T is an interval of the real line, finite or infinite. The basic tool in dealing with the continuous-time parameter process is the sampling which converts a continuous-time parameter process to one discrete in time. The problem is attempted, when the process is not stationary, via a series representation of $\{x(t), t\varepsilon T\}$. This is done in Section 4.2 where it is shown that our criterion of maximizing the Bhattacharyya distance yields an integral equation of Fredholm type to be solved. Section 4.3 contains an explicit expression for the linear discriminant function in the case when the process is covariance stationary ; we have used the Shannon's sampling theorem to deal with the above case.

4.2 SECOND ORDER TIME SERIES

The observed process $\{x(t), t\varepsilon T\}$ is assumed to be continuous in time. Let $T = [o,A]$, where A is a real number (finite). It is also assumed to be a second order process i.e. $EX^2(t) < \infty$ $\forall\, t \,\varepsilon\, [o,A]$.

Let $E_{H_j} X(t) = \mu_j(t)$ (4.2.1)

$Cov_{H_j} (X(t),X(u)) = R_j(t,u)$ (4.2.2)

where $j = 1,2$, t, $u \in [o,A]$.

Assume , $|\delta(t)| = |\mu_1(t) - \mu_2(t)| < \infty$ in $[o,A]$.

Our first step is to reduce the process $\{x(t), t\varepsilon T\}$ to a set of random variables (possibly a countably infinite set)([63]). This is achieved by the method of the series expansion :

$$x(t) = \sum_{n=1}^{\infty} x_n \varphi_n(t)$$ (4.2.3)

in the following sense :

$$\lim_{K \to \infty} E(X(t) - \sum_{n=1}^{K} x_n \varphi_n(t))^2 = 0, \quad 0 \leq t \leq A$$

where

$$x_n = \int_{o}^{A} x(t) \varphi_n(t) dt$$ (4.2.4)

and $\{\varphi_n(t)\}$ is a set of complete orthonormal functions in the interval $[o,A]$ with

$$\int_{o}^{A} \varphi_n(t) \varphi_m(t) dt = \delta_{nm} = \begin{cases} 1 & n = m \\ o & n \neq m \end{cases} ,$$

(see Appendix F).

Let $\alpha(t) \varepsilon L^2(o,A)$ i.e. $\alpha(t)$ is a square integrable function on $[o,A]$. Then the Fourier series of $\alpha(t)$ is

$$\alpha(t) = \sum_{n=1}^{\infty} \alpha_n \varphi_n(t)$$ (4.2.5)

where

$$\alpha_n = \int_{o}^{A} \alpha(t) \varphi_n(t) dt$$ (4.2.6)

Consider K terms in the series (4.2.3), having the coefficients $(x_1, \ldots, x_K)$. Let $\underset{\sim}{x} = (x_1, \ldots, x_K)$.

Then $\underset{\sim}{x} \sim N(\underset{\sim}{\mu}_j, R_j)$, under H_j $(j = 1,2)$,

where
$$\underset{\sim}{\mu}_j = (\mu_{j1}, \ldots, \mu_{jK})$$
$$R_j = ((\text{Cov}_{H_j}(X_i, X_\ell)))_{i, \ell = 1, \ldots, K}.$$

Following the discussions in the previous Chapter, our classification rule would have been:

$$\sum_{n=1}^{K} \alpha_n x_n \overset{H_1}{\underset{H_2}{\gtrless}} c$$

(accept)

But
$$\lim_{K \to \infty} \sum_{n=1}^{K} \alpha_n x_n = \lim_{K \to \infty} \sum_{n=1}^{K} \left(\int_0^A \alpha(t) \varphi_n(t) \, dt \right) x_n$$

$$= \lim_{K \to \infty} \int_0^A \alpha(t) \left(\sum_{n=1}^{K} x_n \varphi_n(t) \right) dt$$

$$= \int_0^A \alpha(t) \, x(t) \, dt,$$

the limit being in quadratic mean.

Hence our linear procedure in the case when the process is continuous in time is given by

$$y \overset{\Delta}{=} \int_0^A \alpha(t) \, x(t) \, dt \overset{H_1}{\underset{H_2}{\gtrless}} c \qquad (4.2.7)$$

(accept)

(For definition of Stochastic integral see Appendix I).

Thus our problem is now to find an optimal $\alpha(t)$, $t\varepsilon[o,A]$ in the sense of maximizing $-\ln P_2(1,2;y)$, y is defined in (4.2.7).

The following theorem states the method of obtaining $\alpha(t)$.

<u>Theorem 4.2.1</u> : The $\alpha(t)$, $t\varepsilon[o,A]$, for which $-\ln P_2(1,2;y)$ is a maximum satisfies a Fredholm integral equation of the first kind

$$\int_o^A R_\theta(t,u)\ \alpha(t)\ dt = \delta(u), \qquad (4.2.8)$$

to be solved iteratively, where θ is given by

$$-\theta = \frac{\left\{\dfrac{\int_o^A \alpha(t)\delta(t)dt}{\int_o^A\int_o^A R(t,u)\alpha(t)\alpha(u)dtdu}\right\}^2 + \dfrac{1}{\int_o^A\int_o^A R_2(t,u)\alpha(t)\alpha(u)dtdu} - \dfrac{2}{\int_o^A\int_o^A R(t,u)\alpha(t)\alpha(u)dtdu}}{\left\{\dfrac{\int_o^A \alpha(t)\delta(t)dt}{\int_o^A\int_o^A R(t,u)\alpha(t)\alpha(u)dtdu}\right\}^2 + \dfrac{1}{\int_o^A\int_o^A R_1(t,u)\alpha(t)\alpha(u)dtdu} - \dfrac{2}{\int_o^A\int_o^A R(t,u)\alpha(t)\alpha(u)dtdu}}$$

$$(4.2.9)$$

and $R_\theta(t,u) \overset{\Delta}{=} R_1(t,u) - \theta R_2(t,u)$,

$\qquad R(t,u) \overset{\Delta}{=} R_1(t,u) + R_2(t,u)$.

The iteration continues until $|\theta^{(i+1)} - \theta^{(i)}| < \varepsilon$, where ε is a pre-assigned number, i denotes the number of iteration.

Proof : For a given u, the function $R_1(t,u)$ can be expanded into a Fourier series in the interval $[o,A]$:

$$R_1(t,u) = \sum_{n=1}^{\infty} \beta_{1n}(u) \, \varphi_n(t), \qquad\qquad (4.2.10)$$

where

$$\beta_{1n}(u) = \int_0^A R_1(t,u) \, \varphi_n(t) \, dt \qquad\qquad (4.2.11)$$

Let

$$R_1(n,m) = \text{Cov}(X_n, X_m) \quad (n,m = 1,\ldots,K)$$

where

$$x_n = \int_0^A x(t) \, \varphi_n(t) \, dt.$$

Thus

$$x_n - \mu_{1n} = \int_0^A (x(t) - \mu_1(t)) \, \varphi_n(t) \, dt,$$

since

$$\mu_1(t) = \sum_{n=1}^{\infty} \mu_{1n} \varphi_n(t), \text{where } \mu_{1n} = \int_0^A \mu_1(t) \, \varphi_n(t) \, dt.$$

Now $R_1(n,m)$ can be written as

$$R_1(n,m) = E(X_n - \mu_{1n})(X_m - \mu_{1m})$$

$$= E \{ \int_0^A (x(t) - \mu_1(t)) \, \varphi_n(t) dt \} \{ \int_0^A (x(t) - \mu_1(t)) \varphi_m(t) dt \}$$

$$= \int_0^A \int_0^A R_1(t,u) \, \varphi_n(t) \, \varphi_m(u) dt \, du$$

$$= \int_0^A (\int_0^A R_1(t,u) \, \varphi_n(t) dt) \, \varphi_m(u) \, du$$

$$= \int_0^A \beta_{1n}(u) \, \varphi_m(u) \, du \quad , \qquad\qquad (4.2.12)$$

by (4.2.11).

If we consider the finite set of random variables $(x_1, \ldots, x_K)$, then the $\underset{\sim}{\alpha}$ vector of the linear procedure satisfies the system of equations

$$R_\theta \underset{\sim}{\alpha} = \underset{\sim}{\delta} \qquad (4.2.13)$$

where

$$-\theta = \frac{\{\dfrac{\alpha' \delta}{\underset{\sim}{\alpha}' (R_1 + R_2) \underset{\sim}{\alpha}}\}^2 + \dfrac{1}{\underset{\sim}{\alpha}' R_2 \underset{\sim}{\alpha}} - \dfrac{2}{\underset{\sim}{\alpha}' (R_1 + R_2) \underset{\sim}{\alpha}}}{\{\dfrac{\underset{\sim}{\alpha}' \underset{\sim}{\delta}}{\underset{\sim}{\alpha}' (R_1 + R_2) \underset{\sim}{\alpha}}\}^2 + \dfrac{1}{\underset{\sim}{\alpha}' R_1 \underset{\sim}{\alpha}} - \dfrac{2}{\underset{\sim}{\alpha}' (R_1 + R_2) \underset{\sim}{\alpha}}}$$

$$(4.2.14)$$

(cf. (3.3.6) and (3.3.7) ,

and $$R_\theta = R_1 - \theta R_2.$$

The nth equation of (4.2.13) is :

$$\sum_{m=1}^{K} R_\theta(n,m) \, \alpha_m = \delta_n$$

which after being multiplied on both sides by $\varphi_n(u)$ and then summing over n (n = 1, ..., K) yields

$$\sum_{n=1}^{K} \sum_{m=1}^{K} R_\theta(n,m) \, \alpha_m \, \varphi_n(u) = \sum_{n=1}^{K} \delta_n \varphi_n(u)$$

Let $K \to \infty$. Then we have,

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} R_\theta(n,m) \, \alpha_m \, \varphi_n(u) = \delta(u)$$

$$\implies \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} R_\theta(m,n) \, \alpha_n \, \varphi_m(u) = \delta(u)$$

$$\implies \sum_{n,m=1}^{\infty} R_\theta(n,m) \left(\int_o^A \alpha(t) \, \varphi_n(t) dt \right) \varphi_m(u) = \delta(u)$$

$$\implies \int_o^A \left(\sum_{n,m} R_\theta(n,m) \, \varphi_n(t) \, \varphi_m(u) \right) \alpha(t) dt = \delta(u),$$

(since $|\delta(u)| < \infty$ in $[o,A]$, $\int$ and Σ can be interchanged),

which can also be written as

$$\int_o^A (\sum_{n,m} R_1(n,m) \; \varphi_n(t) \; \varphi_m(u))\alpha(t)dt \; -\theta \int_o^A (\sum_{n,m} R_2(n,m)\varphi_n(t)\varphi_m(u))\alpha(t)dt = \delta ($$

$$(4.2.15)$$

The first term on the left hand side of (4.2.15) can
be written as

$$\int_o^A (\sum_{n,m} R_1(n,m) \; \varphi_n(t) \; \varphi_m(u))\alpha(t)dt$$

$$= \int_o^A \{ \sum_{n,m} (\int_o^A \beta_{1n}(v)\varphi_m(v)dv) \; \varphi_n(t) \; \varphi_m(u)\} \; \alpha(t)dt \; , \; \text{(by 4.2.12)}$$

$$= \int_o^A [\sum_m \{ \int_o^A (\sum_{n=1}^\infty \beta_{1n}(v) \; \varphi_n(t)) \varphi_m(v)dv \}]\alpha(t)dt \; \varphi_m(u)$$

$$= \int_o^A \{ \sum_{m=1}^\infty (\int_o^A R_1(t,v) \; \varphi_m(v)dv) \; \varphi_m(u) \} \alpha(t)dt \; , \; \text{(by 4.2.10)}$$

$$= \int_o^A (\sum_{m=1}^\infty \beta_{1m}(t) \; \varphi_m(u))\alpha(t)dt \quad , \; \text{(by 4.2.11)}$$

$$= \int_o^A R_1(u,t)\alpha(t)dt \quad , \quad \text{(by 4.2.10)}$$

$$= \int_o^A R_1(t,u)\alpha(t)dt \quad\quad\quad (4.2.16)$$

It is to be noted that what we have actually shown above is

$$\sum_{n,m=1}^\infty R_1(n,m) \; \varphi_n(t) \; \varphi_m(u) = R_1(t,u) \quad\quad (4.2.17)$$

By a similar argument, the second term on the left hand side

of (4.2.15) can be put in the form

$$\theta \int_0^A R_2(t,u)\alpha(t)dt \tag{4.2.18}$$

Using (4.2.16) and (4.2.18), (4.2.15) reduces to (4.2.8):

$$\int_0^A R_9(t,u)\alpha(t)dt = \delta(u)$$

where θ is given by (using 4.2.14),

$$-\theta = \lim_{K \to \infty} \frac{\left\{\dfrac{\sum\limits_{n=1}^{K} \alpha_n\delta_n}{\sum\limits_{n,m}^{K} R(n,m)\alpha_n\alpha_m}\right\}^2 + \dfrac{1}{\sum\limits_{n,m}^{K} R_2(n,m)\alpha_n\alpha_m} - \dfrac{2}{\sum\limits_{n,m}^{K} R(n,m)\alpha_n\alpha_m}}{\left\{\dfrac{\sum\limits_{n=1}^{K} \alpha_n\delta_n}{\sum\limits_{n,m=1}^{K} R(n,m)\alpha_n\alpha_m}\right\}^2 + \dfrac{1}{\sum\limits_{n,m}^{K} R_2(n,m)\alpha_n\alpha_m} - \dfrac{2}{\sum\limits_{n,m}^{K} R(n,m)\alpha_n\alpha_m}}$$

$\qquad$ = right hand side of (4.2.9),

which follows from the following two facts :

$$1) \lim_{K \to \infty} \sum_{n=1}^{K} \alpha_n\delta_n = \lim_{K \to \infty} \sum_{n=1}^{K} \left(\int_0^A \alpha(t)\,\varphi_n(t)dt\right)\delta_n$$

$$= \sum_{n=1}^{\infty} \left(\int_0^A \alpha(t)\,\varphi_n(t)dt\right)\delta_n$$

$$= \int_0^A \left(\sum_{n=1}^{\infty} \delta_n\,\varphi_n(t)\right)\alpha(t)dt$$

$$= \int_0^A \alpha(t)\delta(t)dt.$$

2) (i) $\lim\limits_{K\to\infty} \underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha} = \lim\limits_{K\to\infty} \overset{K}{\underset{n,m=1}{\Sigma}} R_1(n,m)\ \alpha_n\alpha_m$

$$= \overset{\infty}{\underset{n,m=1}{\Sigma}} R_1(n,m)(\overset{A}{\underset{o}{\int}}\alpha(t)\varphi_n(t)dt)(\overset{A}{\int}\alpha(u)\varphi_m(u)du)$$

$$= \overset{A}{\underset{o}{\int}}\overset{A}{\underset{o}{\int}}(\overset{\infty}{\underset{n,m=1}{\Sigma}} R_1(n,m)\varphi_n(t)\varphi_m(u))\alpha(t)\alpha(u)dtdu$$

$$= \overset{A}{\underset{o}{\int}}\overset{A}{\underset{o}{\int}} R_1(t,u)\alpha(t)\alpha(u)dtdu,$$

using (4.2.17).

(ii) By similar arguments,

$$\lim\limits_{K\to\infty} \underset{\sim}{\alpha}'R_2\underset{\sim}{\alpha} = \overset{A}{\underset{o}{\int}}\overset{A}{\underset{o}{\int}} R_2(t,u)\alpha(t)\alpha(u)dtdu.$$

Hence the theorem is proved.

<u>Remark 4.2.1</u> : It may happen that the integral equation (4.2.8),

for a given θ , does not have a solution. This can be seen as

follows ([62]). To this end the following theorem may be

useful.

<u>Theorem</u> (<u>Hilbert-Schmidt theorem</u>)

If $\delta(u)$ can be written in the form

$$\delta(u) = \overset{A}{\underset{o}{\int}} R_\theta(t,u)\alpha(t)dt \qquad\qquad (4.2.19)$$

where $R_\theta(t,u)$ is a symmetric L_2-kernel (i.e. $\overset{A}{\underset{o}{\int}}\overset{A}{\underset{o}{\int}} R_\theta^2(t,u)dtdu < \infty$

and $R_\theta(t,u) = R_\theta(u,t)$) and $\overset{A}{\underset{o}{\int}}\alpha^2(t)dt < \infty$, then we can write

$$\delta(u) = \overset{\infty}{\underset{h=1}{\Sigma}} a_h\ \Psi_h(u), \qquad\qquad (4.2.20)$$

where

$$a_h = \int_0^A \delta(u)\, \Psi_h(u)\, du \quad (h = 1,2,\dots) \qquad (4.2.21)$$

and $\{\Psi_h(x)\}$ is the orthonormal system of eigenfunctions of

$R_\theta(t,u)$ satisfying

$$\int_0^A R_\theta(t,u)\, \Psi_h(u)\, du = \frac{\Psi_h(t)}{\lambda_h}\ , \qquad (4.2.22)$$

λ_h's are called the eigenvalues of the kernel.

Let

$$\alpha(t) = \sum_{h=1}^{\infty} c_h\, \Psi_h(t)\ , \qquad (4.2.23)$$

where

$$c_h = \int_0^A \alpha(t)\, \Psi_h(t)\, dt. \qquad (4.2.24)$$

Then from (4.2.19), we get

$$\delta(u) = \int_0^A R_\theta(t,u)\Big(\sum_h c_h\, \Psi_h(t)\Big)\, dt$$

$$= \sum_h \Big(\int_0^A R_\theta(t,u)\, \Psi_h(t)\, dt\Big)\, c_h$$

$$= \sum_{h=1}^{\infty} \Big(\frac{c_h}{\lambda_h}\Big)\, \Psi_h(u),$$

using (4.2.22).

Thus, $a_h = \dfrac{c_h}{\lambda_h} \quad (h = 1,2,\dots).$

$$\Longrightarrow c_h = a_h\, \lambda_h$$

Hence (4.2.23) can be rewritten as

$$\alpha(t) = \sum_{h=1}^{\infty} a_h \lambda_h \Psi_h(t), \qquad (4.2.25)$$

which is the solution of our equation, provided $\int_o^A \alpha^2(t)dt < \infty$, which is the same as saying that

$$\sum_{h=1}^{\infty} a_h^2 \lambda_h^2 < \infty. \qquad (4.2.26)$$

If the infinite series in (4.2.26) diverges, then our equation does <u>not</u> have a solution in L_2-class.

4.3 WHEN THE PROCESS IS STATIONARY

We divide the analysis into two cases depending on the nature of the parameter set of the process $\{x(t), t\epsilon T\}$.

<u>Case 1</u> When T is a finite interval $[o,A]$.

We shall discuss two types of kernels in which straightforward procedures for solving (4.2.8) are available for a given θ.

<u>Type A</u> (Rational Kernels)

Let $R_\theta(t,u) = R_\theta(t-u)$, $t,u\epsilon[o,A]$.

Let the Fourier transform of this be

$$S_\theta(\omega) = \int_{-\infty}^{\infty} e^{i\omega\tau} R_\theta(\tau)d\tau \stackrel{\Delta}{=} \frac{N(\omega^2)}{D(\omega^2)} \qquad (4.3.1)$$

This is the ratio of two polynomials in ω^2. The kernels whose transforms satisfy (4.3.1) are called <u>rational kernels</u>. The basic technique is to find a differential equation corresponding to the integral equation. Because of the form

of the kernel, this will be a differential equation with constant coefficients whose solution can be readily obtained.

Let $\delta_D(t)$ be the Dirac delta function, so that

$$\int_{-\infty}^{\infty} \delta_D(t) \ e^{i\omega t} \ dt = 1.$$

(see, Appendix G)

$$\Longrightarrow \delta_D(t-u) = \int_{-\infty}^{\infty} e^{iw(t-u)} \ \frac{d\omega}{2\pi} \qquad (4.3.2)$$

Differentiating (Appendix G) this with respect to t gives

$$p\delta_D(t-u) = \int_{-\infty}^{\infty} i\omega e^{i\omega(t-u)} \ \frac{d\omega}{2\pi} \qquad (4.3.3)$$

where $\qquad p \overset{\Delta}{=} \dfrac{d}{dt} \ .$

More generally,

$$N(-p^2)\delta_D(t-u) = \int_{-\infty}^{\infty} N(\omega^2) \ e^{i\omega(t-u)} \ \frac{d\omega}{2\pi} \quad (4.3.4)$$

In an analogous fashion,

$$D(-p^2)R_\theta(t-u) = \int_{-\infty}^{\infty} D(\omega^2)S_\theta(\omega) \ e^{i\omega(t-u)} \ \frac{d\omega}{2\pi}$$

$$= \int_{-\infty}^{\infty} N(\omega^2) \ e^{i\omega(t-u)} \ \frac{d\omega}{2\pi} \qquad (4.3.5)$$

Comparing (4.3.4) and (4.3.5), we get

$$N(-p^2)\delta_D(t-u) = D(-p^2)R_\theta(t-u) \qquad (4.3.6)$$

Operating on both sides of (4.2.8) with $D(-p^2)$, we obtain,

$$D(-p^2)\delta(t) \qquad = \int_{0}^{A} D(-p^2)R_\theta(t-u)\alpha(u)du$$

$$= \int_0^A N(-p^2)\, \delta_D(t-u)\, \alpha(u)du, \quad \text{(using 4.3.6)}$$

$$= N(-p^2)\, \alpha(t) \quad , \; 0 \le t \le A$$

i.e. the differential eqn. of our interest is

$$D(-p^2)\delta(t) = N(-p^2)\,\alpha(t) \qquad\qquad (4.3.7)$$

Some specific examples are given in VanTrees ([63])

for illustration.

The following simple example illustrates the technique.

Example 4.3.1 .

Take $R_1(t-u) = R_2(t-u) = R_2(\tau)$, $(-\infty < \tau < \infty)$
$$= e^{-|\tau|}$$

or $$S_\theta(\omega) = \frac{4}{\omega^2+1}.$$

Therefore, $N(\omega^2) = 4$
$$D(\omega^2) = (1+\omega^2),$$

and the differential equation (4.3.7) is

$$-\frac{d^2\delta(t)}{dt^2} + \delta(t) = 4\alpha(t) , \quad 0 < t < A$$

whence $\alpha(t)$ can be readily obtained.

Type B . Let us consider the problem,

$$x(t) = \begin{cases} \mu_1(t) + n_1(t) & \text{under } H_1 \\ \mu_2(t) + n_2(t) & , \text{under } H_2 \end{cases}, \quad t\epsilon[o,A].$$

Suppose $n_1(t)$ contains only the white noise,

i.e. $E_{H_1} n_1(t) n_1(u) = \delta_D(t-u).$

Then (4.2.8) reduces to

$$\int_0^A \delta_D(t-u)\alpha(t)dt - \theta \int_0^A R_2(t,u)\alpha(t)dt = \delta(u)$$

or,
$$\alpha(u) - \theta\int_0^A R_2(t,u)\alpha(t)dt = \delta(u) \qquad (4.3.8)$$

which is known as the Fredholm integral equation of the
second kind.

Assume $R_2(t,u)$ is a L_2 kernel and $\int_0^A R_2^2(t,u)du < \infty$.

We note that $\alpha(u) - \delta(u)$ has an integral representation
of the form (4.2.19). Hence by Hilbert-Schmidt theorem,
it can be represented by an absolutely and uniformly convergent
series of the type :

$$\alpha(u) - \delta(u) = \sum_{h=1}^{\infty} d_h \tilde{\varphi}_h(u), \qquad (4.3.9)$$

where $\tilde{\varphi}_h(u)$'s are given by

$$\int_0^A R_2(t,u)\,\tilde{\varphi}_h(u)du = \frac{\tilde{\varphi}_h(t)}{\nu_h} \qquad (4.3.10)$$

and
$$d_h = \int_0^A (\alpha(u) - \delta(u))\tilde{\varphi}_h(u)\,du$$

$$= e_h - \xi_h$$

with
$$\xi_h = \int_0^A \delta(u)\,\tilde{\varphi}_h(u)du, \quad e_h = \int_0^A \alpha(u)\,\tilde{\varphi}_h(u)\,du$$

$$(4.3.11)$$

But
$$d_h = \theta \int_0^A (\int_0^A R_2(t,u)\alpha(t)dt)\,\tilde{\varphi}_h(u)du$$

$$= \theta \int_o^A \left(\int_o^A R_2(t,u)\, \tilde{\varphi}_h(u)\, du \right) \alpha(t)\, dt$$

(since $d_h < \infty$ and $\int R_2(t,u)\, \tilde{\varphi}_h(u)\, du < \infty$, interchanging the order of integration is permitted)

$$= \theta \int_o^A \frac{\tilde{\varphi}_h(t)}{\nu_h}\, \alpha(t)\, dt \quad , \text{ (by 4.2.35)}$$

$$= \theta\, \frac{e_h}{\nu_h} \quad , \text{ (by 4.2.36)} \tag{4.3.12}$$

$$\Longrightarrow\ e_h - \xi_h = \theta\, \frac{e_h}{\nu_h}$$

$$\Longrightarrow\ \xi_h = e_h\left(1 - \frac{\theta}{\nu_h}\right)$$

$$\Longrightarrow\ e_h = \frac{\nu_h}{\nu_h - \theta}\, \xi_h \tag{4.3.13}$$

Using (4.3.13), $\quad d_h = \theta\, \dfrac{\nu_h}{\nu_h - \theta}\, \xi_h\, \dfrac{1}{\nu_h}$

$$= \theta\, \frac{\xi}{\nu_h - \theta} \tag{4.3.14}$$

Hence from (4.3.9),

$$\alpha(u) = \delta(u) + \theta \sum_{h=1}^{\infty} \frac{\xi_h}{\nu_h - \theta}\, \tilde{\varphi}_h(u)$$

$$= \delta(u) + \theta \sum_{h=1}^{\infty} \int_o^A \frac{\tilde{\varphi}_h(t)\, \tilde{\varphi}_h(u)}{\nu_h - \theta}\, \delta(t)\, dt \tag{4.3.15}$$

provided θ is <u>not</u> an eigenvalue of $R_2(t,u)$.

Remark 4.3.1 : When $R_1(t,u) = R_2(t,u)$, (4.2.8) reduces to

$$2 \int_0^A R_1(t,u)\alpha(t)dt = \delta(u),\qquad (4.3.16)$$

because $\theta = -1$.

The equation (4.3.16) is studied widely in ([63]). Let us consider an important special case. Suppose, $n_1(t)$ and $n_2(t)$ both are white.

Since $\delta_D(t-u)$ is a p.d. function, by Mercer's theorem,

$$\delta_D(t-u) = \sum_{h=1}^{\infty} (\tilde{\varphi}_h(t) \tilde{\varphi}_h(u))\qquad (4.3.17)$$

where $\tilde{\varphi}_h(t)$'s are eigen functions of $\delta_D(t-u)$.

Using (4.3.17), from (4.3.15) we obtain,

$$\begin{aligned}
\alpha(u) &= \delta(u) - \frac{1}{2} \int_0^A (\sum_1^{\infty} \tilde{\varphi}_h(t) \tilde{\varphi}_h(u)) \delta(t)dt \\
&= \delta(u) - \frac{1}{2} \int_0^A \delta_D(t-u)\delta(t)dt \\
&= \delta(u) - \frac{1}{2} \delta(u) \\
&= \frac{1}{2} \delta(u),
\end{aligned}$$

which is a well-known result (see [63]).

Case 2. When T is an infinite interval $[o, \infty)$

We make the following assumptions analogous to those required for discrete-time parameter process :

$\underline{C1}$. The difference of the mean functions $\delta(t)$ satisfies

i) $\sup_t |\delta(t)| < \infty$

ii) $\lim_{A \to \infty} \frac{1}{A} \int_0^{A-1-|\tau|} \delta(t)\delta(t+\tau)dt = \int_{-\pi}^{\pi} e^{i\lambda\tau} \frac{dM(\lambda)}{2\pi}$, $0 < \tau < \infty$.

where $M(\lambda)$ is as given in Chapter III.

C2. The covariance functions $R_j(v)$ satisfy

$$\int_0^\infty |R_j(v)| \; |v|^{1+\beta} \, dv < \infty, \quad \text{where } 0 < \beta < 1, \quad (j = 1,2).$$

C3. The spectral density of the covariance stationary process $\{x(t), \; 0 < t < \infty\}$ is zero outside the frequency interval $-\pi \le \lambda \le \pi$.

Then we have a theorem , which gives an explicit expression for $\alpha(t)$.

Theorem 4.3.1: Assume C1 to C3 of the above are satisfied while the underlying process is covariance stationary. Then the optimal $\alpha(t)$ is given by

$$\alpha(t) = \int_{-\infty}^\infty \frac{D(\lambda)}{S(\lambda)} \; e^{-i\lambda t} \frac{d\lambda}{2\pi}, \quad (0 < t < \infty) \qquad (4.3.18)$$

where

$$D(\lambda) = \int_0^\infty \delta(u) \, e^{i\lambda u} \, du \qquad (4.3.19)$$

and

$$S(\lambda) = \int_0^\infty R(v) \, e^{i\lambda v} \, dv \qquad (4.3.20)$$

We need the following lemmas to prove the theorem.

Lemma 4.3.1. Let $\mathbb{R} = (-\infty, +\infty)$. Assume that $F(x)$ and $G(x)$ are Lebesgue integrable on $\mathbb{R}$ and that at least one of F and G is continuous and bounded on $\mathbb{R}$. Let

$$H(x) = \int_{-\infty}^\infty F(t)G(x-t) \, dt.$$

Then for every real λ , we have,

$$\int_{-\infty}^{\infty} H(x)\ e^{i\lambda x}\ dx = (\int_{-\infty}^{\infty} F(y)\ e^{i\lambda y}\ dy)(\int_{-\infty}^{\infty} G(z)\ e^{i\lambda z}\ dz)$$

$$(4.3.21)$$

Proof : See Apostol ([6], pp. 329-331).

Lemma 4.3.2. If for some Λ (real number), the spectral density of a covariance stationary process $\{ X(t),\ o < t < \infty \}$ is zero outside of the frequency interval $-\Lambda \leq \lambda \leq \Lambda$, then the process can be exactly reconstructed from its values at the time points

$$\frac{\pi k}{\Lambda}\ (k = o, 1, 2, \ldots)$$

More precisely,

$$X(t) = \sum_{k=o}^{\infty} \frac{\sin \Lambda\ (t - \frac{\pi k}{\Lambda})}{\Lambda\ (t - \frac{\pi k}{\Lambda})}\ x\ (\frac{\pi k}{\Lambda}),\ (o < t < \infty)$$

$$(4.3.22)$$

Proof : This result is known as the Shannon's sampling theorem, the proof of which can be found in Koopmans ([29]).

Proof of the Theorem 4.3.1 :

By Lemma 4.3.2, we have

$$x(t) = \sum_{n=o}^{\infty} \varphi_n(t)\ x(n),\qquad (4.3.23)$$

where $\qquad \varphi_n(t) = \dfrac{\sin \pi (t - n)}{\pi\ (t - n)}$.

Thus our sampled process is $\{ x(k),\ k = o, 1, 2, \ldots \}$; which is also covariance stationary and satisfies all the assumptions

A1 to A4 laid down in the Section 3.5 of Chapter III. By an application of the theory therein, if we consider K' terms in the series of (4.3.23), we have

$$(R_1 + R_2) \underset{\sim}{\alpha} = \underset{\sim}{\delta} .$$

Now we let $K' \longrightarrow \infty$. Then by a similar argument as used in Section 4.2, we get,

$$\int_0^\infty R(t-u) \, \alpha(t) \, dt = \delta(u) \qquad (4.3.24)$$

where $R(t-u) = R_1(t-u) + R_2(t-u)$.

Let $A(\lambda) = \int_0^\infty \alpha(t) \, e^{i\lambda t} \, d\lambda$ i.e. $A(\lambda)$ is the Fourier transform of $\alpha(t)$.

Then, using lemma 4.3.1, we obtain from (4.3.24),

$$S(\lambda) \cdot A(\lambda) = D(\lambda), \qquad (4.3.25)$$

where $S(\lambda)$, $D(\lambda)$ are defined in (4.3.19) and (4.3.20).

(4.3.25) implies $A(\lambda) = \dfrac{D(\lambda)}{S(\lambda)}$

$$\Longrightarrow \alpha(t) = \int_{-\infty}^{\infty} \frac{D(\lambda)}{S(\lambda)} \, e^{-i\lambda t} \, \frac{d\lambda}{2\pi} ,$$

which is the same as (4.3.18) . This completes the proof.

4.4 CONCLUSION

The linear discriminant function is

$$y \overset{\Delta}{=} \int_0^A \alpha(t) \, x(t) \, dt.$$

Our problem was to find an $\alpha(t)$ such that $-\ln\rho_2(1,2;y)$ given by

$$-\ln\rho_2(1,2;y) = \frac{1}{4} \ln(\int_o^A \int_o^A \alpha(t)R_1(t,u)\alpha(u)dtdu)$$

$$+ \frac{1}{4} \ln(\int_o^A \int_o^A \alpha(t)R_2(t,u)\alpha(u)dtdu)$$

$$- \frac{1}{2} \ln(\frac{1}{2} \{ \int_o^A \int_o^A \alpha(t)R(t,u)\alpha(u)dtdu \})$$

$$+ \frac{1}{4} \frac{(\int_o^A \alpha(t)\delta(t)dt)^2}{(\int_o^A \int_o^A \alpha(t)R(t,u)\alpha(u)dtdu)}$$

is a maximum . We attempted the problem through the sampling technique. We noted that no compact form of $\alpha(t)$ is available except the case when our basic process is covariance stationary with the parameter set infinite.

CHAPTER V

DESIGN OF SIGNALS

5.1 INTRODUCTION

This chapter treats the problem of selecting a signal
embedded in normally distributed additive noise with zero
mean and possibly unequal covariance matrices. We state the
problem mathematically in Section 5.2. The various methods
of obtaining an optimal signal are discussed in Section 5.3
while Section 5.4 contains some numerical results.

5.2 MATHEMATICAL STATEMENT OF THE PROBLEM OF SIGNAL SELECTION

A fairly general mathematical model of the time series
(discrete) $\{X(t), t \varepsilon T\}$ can be written as

$$X(t) = \mu(t) + n(t),$$

where $\mu(t)$ is a completely deterministic process and $n(t)$
is a stochastic process. They are sometimes called ''signal''
and ''noise'' respectively.

Let
$$X(t) = \begin{cases} \mu(t) + n_1(t), & \text{under } H_1 \\ n_2(t), & \text{under } H_2 \end{cases},$$

where $n_j(t)$'s $(j = 1,2)$ are normal processes.

Let $\underset{\sim}{x}$ denote n observations on $X(t)$; i.e.,

$$\underset{\sim}{x} = (X(o),\dots,X(n-1))$$
$$\underset{\sim}{\mu} = (\mu(o),\dots,\mu(n-1))$$
$$\underset{\sim}{n_j} = (n_j(o),\dots,n_j(n-1))$$

where let $\underset{\sim}{n}_j \sim N_n(\underset{\sim}{o}, R_j)$ under $H_j (j = 1,2)$

or equivalently, $\underset{\sim}{x} \sim N_n(\mu, R_1)$ under H_1

and $\underset{\sim}{x} \sim N_n(\underset{\sim}{o}, R_2)$ under H_2.

Our object (from the view point of classification) is to minimize the Bayes risk, or, if we attack equal costs to the two types of errors, to minimize the total probability of misclassification. For a given $\underset{\sim}{\mu}$, the signal vector, this probability will also be a function of $\underset{\sim}{\mu}$ and one naturally asks : which is the signal vector that minimizes the total probability of error resulting from the use of the Bayes optimal classification rule subject to the restriction $\underset{\sim}{\mu}'\underset{\sim}{\mu} = 1$? Unfortunately, it seems difficult to carry out the direct minimization involved. Hence we resort to some signal selection criteria ([22,26,48]) that may be weakar than the error probability but are easier to evaluate and manipulate.

Our first step is to classify $\underset{\sim}{x}$ into H_j $(j = 1,2)$ optimally is some sense. This we do via linear procedures considered in earlier chapters. But our criterion is now of minimizing the total probability of error resulting from the setting up of the logical requirement that each probability of misclassificat is the same when no knowledge is available about the a-priori probability of the hypotheses. If we denote the total probability of error by $Pr(\varepsilon)$, then our problem is :

$$\min_{\substack{\mu \\ \mu' \mu = 1}} \quad \min_{\alpha} \quad Pr(\varepsilon) \qquad (5.2.1)$$

5.3 METHODS OF SIGNAL SELECTION

Method 1 : If we apply the classification scheme

$$\alpha' x \underset{H_2}{\overset{H_1}{\gtrless}} c$$

(accept)

then, as we have seen in Chapter III, the two types of errors result which are given by

$$e_1 = 1 - \Phi(y_1), \text{ where } y_1 = \frac{\alpha' \mu - c}{(\alpha' R_1 \alpha)^{1/2}} \qquad (5.3.1)$$

and $e_2 = 1 - \Phi(y_2)$, where $y_2 = \dfrac{c}{(\alpha' R_2 \alpha)^{1/2}} \qquad (5.3.2)$

Set $e_1 = e_2$, which is equivalent to saying $y_1 = y_2$, since $\Phi(x)$ is montone in x.

This implies

$$\frac{\alpha' \mu - c}{(\alpha' R_1 \alpha)^{1/2}} = \frac{c}{(\alpha' R_2 \alpha)^{1/2}}$$

$$\Longrightarrow \quad c = \frac{\alpha' \mu (\alpha' R_2 \alpha)^{1/2}}{(\alpha' R_1 \alpha)^{1/2} + (\alpha' R_2 \alpha)^{1/2}} \qquad (5.3.3)$$

Putting this value of c in the expression for y_1 in (5.3.1), we obtain

$$Y_1 = \frac{\alpha' \mu}{(\alpha' R_1 \alpha)^{1/2} + (\alpha' R_2 \alpha)^{1/2}} \tag{5.3.4}$$

which is positive ; since the case $Y_1 < 0$ is of no interest, because the minimum $Pr(\varepsilon)$ achievable in this case is $\frac{1}{2}$.

Now, if ω_j is the a-priori probability of H_j $(j = 1,2)$, then the total probability of error is given by

$$Pr(\varepsilon) = \omega_1 e_1 + \omega_2 e_2$$
$$= e_1$$

since $e_1 = e_2$.

Since $\Phi(Y_1)$ is monotone in Y_1 , our problem (5.2.1) can be reformulated as

$$\max_{\substack{\mu \\ \mu' \mu = 1}} \max_{\alpha} \frac{\alpha' \mu}{(\alpha' R_1 \alpha)^{1/2} + (\alpha' R_2 \alpha)^{1/2}} , \tag{5.3.5}$$

which can also be written as

$$Y_1^{**} \overset{\Delta}{=} \max_{\alpha} \max_{\substack{\mu \\ \mu' \mu = 1}} \frac{\alpha' \mu}{(\alpha' R_1 \alpha)^{1/2} + (\alpha' R_2 \alpha)^{1/2}} \tag{5.3.6}$$

This is due to the assumption that α belongs to a n-cell in $\mathbb{R}^n$, the n-dimensional Euclidean space.

Now, $\quad Y_1^{**} = \max_{\alpha} \dfrac{(\alpha' \alpha)^{1/2}}{(\alpha' R_1 \alpha)^{1/2} + (\alpha' R_2 \alpha)^{1/2}} , \tag{5.3.7}$

since by Cauchy-Schwarz inequality,

$$(\alpha' \mu)^2 \leqq (\alpha' \alpha)(\mu' \mu),$$

where the equality is obtained when $\underset{\sim}{\mu} = \underset{\sim}{\alpha}$.   (5.3.8)

Since R_1 and R_2 are p.d. matrices, they can be simultaneously diagonalized (see Anderson [3], pp. 341). There is a non-singul matrix P such that

$$R_1 = P' \Lambda P$$
$$\text{and} \quad R_2 = P' P \qquad (5.2.9)$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n (> 0)$ are the roots of

$$|R_1 - \lambda R_2| = 0. \qquad (5.3.10)$$

Thus we can reduce (5.3.7) to a simpler form :

$$y_1^{**} = \max_{\substack{\underset{\sim}{\alpha} \\ \underset{\sim}{\alpha}'\underset{\sim}{\alpha} = 1}} \frac{1}{(\underset{\sim}{\alpha}' P' \Lambda P \underset{\sim}{\alpha})^{1/2} + (\underset{\sim}{\alpha}' P' P \underset{\sim}{\alpha})^{1/2}}$$

$$= \max_{\substack{\underset{\sim}{\beta} \\ \underset{\sim}{\beta}' Q \underset{\sim}{\beta} = 1}} \frac{1}{(\underset{\sim}{\beta}' \Lambda \underset{\sim}{\beta})^{1/2} + (\underset{\sim}{\beta}' \underset{\sim}{\beta})^{1/2}}$$

$$= \min_{\substack{\underset{\sim}{\beta} \\ \underset{\sim}{\beta}' Q \underset{\sim}{\beta} = 1}} \left((\underset{\sim}{\beta}' \Lambda \underset{\sim}{\beta})^{1/2} + (\underset{\sim}{\beta}' \underset{\sim}{\beta})^{1/2} \right), \quad (5.3.11)$$

where $Q = (PP')^{-1}$, and $\underset{\sim}{\beta} = P\underset{\sim}{\alpha}$.

We can try to solve (5.3.11) by Lagrange's method.

Let $L = (\underset{\sim}{\beta}' \Lambda \underset{\sim}{\beta})^{1/2} + (\underset{\sim}{\beta}' \underset{\sim}{\beta})^{1/2} + \gamma(\underset{\sim}{\beta}' Q \underset{\sim}{\beta} - 1)$, where γ is the Lagrange's multiplier.

(iii) Form $F = E'R_1E$

(iv) Find an orthogonal matrix H which diagonalizes F,

(v) Form EH

(vi) Then $P = (EH)^{-1}$.

Remark 5.3.2 : It seems that (5.3.12) does not admit of any analytical solution. Of course, one can solve it numerically. However we do not know whether the iteration involved will converge all the time. Even if it does, it is not certain that the solution obtained would give a minimum. Let us give an example.

$$
\text{Take} \quad R_1 = \begin{bmatrix} .12 & .15 \\ .15 & .525 \end{bmatrix} , \quad R_2 = \begin{bmatrix} 1.04 & .7 \\ .7 & 1.25 \end{bmatrix} .
$$

$$
\text{Then} \quad P = \begin{bmatrix} 1 & .5 \\ .2 & 1 \end{bmatrix} , \quad \Lambda = \begin{bmatrix} .1 & 0 \\ 0 & .5 \end{bmatrix} .
$$

A solution of (5.3.12) is (0.9543,0.9543) ; but one can show this does not give the minimum sought for.

This difficulty in carrying out the above procedure leads one to search for some signal selection criteria that may be weaker than the error probability but are easier to evaluate and manipulate. A possible way out is to resort to some bound on the total probability of error and minimize that bound. Before this issue is taken up in Method 2 , we want to make some remarks.

<u>Remark 5.3.3</u> : (i) Let R_2 be a scalar matrix i.e. $R_2 = \nu I$, $\nu >$
Then from (5.3.7), we have,

$$y_1^{**} = \max_{\underset{\sim}{\alpha}} \frac{(\underset{\sim}{\alpha}'\underset{\sim}{\alpha})^{1/2}}{(\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha})^{1/2}+(\nu\underset{\sim}{\alpha}'\underset{\sim}{\alpha})^{1/2}}$$

$$= \max_{\underset{\sim}{\alpha}} \frac{1}{(\frac{\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'\underset{\sim}{\alpha}})^{1/2}+(\nu)^{1/2}}$$

$$= \frac{1}{(\lambda_{min}(R_1))^{1/2} + (\nu)^{1/2}} , \qquad (5.3.1$$

since $\min\limits_{\underset{\sim}{\alpha}} \dfrac{\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'\underset{\sim}{\alpha}} = \lambda_{min}(R_1)$ (see Rao ([57])).

The maximum is attained for the $\underset{\sim}{\alpha}_*$ which is an eigen vector
corresponding to the minimum eigen value of R_1.

(ii) Let $R_1 = R_2$

Then (5.3.5) reduces to

$$\max_{\substack{\underset{\sim}{\mu} \\ \underset{\sim}{\mu}'\underset{\sim}{\mu}=1}} \max_{\underset{\sim}{\alpha}} \frac{\underset{\sim}{\alpha}'\underset{\sim}{\mu}}{2(\underset{\sim}{\alpha}'R_1\underset{\sim}{\alpha})^{1/2}}$$

$$= \max_{\substack{\underset{\sim}{\mu} \\ \underset{\sim}{\mu}'\underset{\sim}{\mu}=1}} \frac{(\underset{\sim}{\mu}'R_1^{-1}\underset{\sim}{\mu})^{1/2}}{2} , \quad (\text{see Rao } ([57]))$$

$$= \frac{1}{2} (\lambda_{max}(R_1^{-1}))^{1/2}$$

the maximum being attained for $\underset{\sim}{\alpha}_* = R_1^{-1}\underset{\sim}{\mu}_*$, and $\underset{\sim}{\mu}_*$ is the

eigen vector corresponding to the minimum eigen value of R_1. This is a well known result (see [48]).

Method 2 : The method is described in the following theorem.

Theorem 5.3.1 : Our signal selection criterion is

$$\max_{\substack{\mu \\ \mu'\mu = 1}} \quad \max_{\alpha} \quad \frac{\alpha'\mu}{(\alpha' R_j \alpha)^{1/2}} \quad (j = 1 \text{ or } 2) \tag{5.3.14}$$

where the optimal α and μ are given by :

$$\alpha_* = R_1^{-1}\mu_* \quad \text{or} \quad R_2^{-1}\mu_*$$

and μ_* is an eigen vector corresponding to the $\lambda_{min}(R_1)$ or $\lambda_{min}(R_2)$ according as

$$\frac{(\lambda_{max}(R_1^{-1}))^{1/2}}{(\lambda_{max}(R_1^{-1}R_2))^{1/2}+1} \quad \gtrless \quad \frac{(\lambda_{max}(R_2^{-1}))^{1/2}}{(\lambda_{max}(R_2^{-1}R_1))^{1/2}+1} \tag{5.3.15}$$

Proof : We have from (5.3.4)

$$y_1 = \frac{\alpha'\mu}{(\alpha' R_1 \alpha)^{1/2} + (\alpha' R_2 \alpha)^{1/2}} \,,$$

which we can write as

$$y_1 = \frac{\alpha'\mu/(\alpha' R_2 \alpha)^{1/2}}{\left(\dfrac{\alpha' R_1 \alpha}{\alpha' R_2 \alpha}\right)^{1/2} + 1} \tag{5.3.16}$$

$$\geq \frac{\alpha'\mu/(\alpha' R_2 \alpha)^{1/2}}{(\lambda_{max}(R_2^{-1}R_1))^{1/2}+1} \,,$$

see Rao ([57]).

Thus our signal selection criterion is

$$B_1^{**} \overset{\Delta}{=} \max_{\substack{\mu \\ \mu'\mu = 1}} \max_{\alpha} \frac{\alpha' \mu}{(\alpha' R_2 \alpha)^{1/2}} \qquad (5.3.17)$$

Now, $B_1^{**} = \max_{\substack{\mu \\ \mu'\mu = 1}} (\mu' R_2^{-1} \mu)^{1/2}$

$$= (\lambda_{max}(R_2^{-1}))^{1/2}.$$

The maximum is attained at $\alpha_* = R_2^{-1}\mu_*$, where the optimal μ is denoted by μ_* which is an eigen vector corresponding to the $\lambda_{min}(R_2)$. Consequently,

$$Y_1^{**} \geq \frac{(\lambda_{max}(R_2^{-1}))^{1/2}}{(\lambda_{max}(R_2^{-1}R_1))^{1/2}+1}.$$

In a similar manner we get by interchanging R_1 and R_2 in the above,

$$Y_1^{**} \geq \frac{(\lambda_{max}(R_1^{-1}))^{1/2}}{(\lambda_{max}(R_1^{-1}R_2))^{1/2}+1}$$

Thus, $Y_1^{**} \geq \max \{ \frac{(\lambda_{max}(R_1^{-1}))^{1/2}}{(\lambda_{max}(R_1^{-1}R_2))^{1/2}+1}, \frac{(\lambda_{max}(R_2^{-1}))^{1/2}}{(\lambda_{max}(R_2^{-1}R_1))^{1/2}+1} \} \overset{\Delta}{=} B^{**}$ (5.3.18)

This completes the proof.

Remark 5.3.4 : If the proposed criterion is $\min_{e_1=k_1 e_2} Pr(\varepsilon)$ then it can be easily shown that the optimum signal is the same as that stated in Theorem 5.3.1.

5.4. NUMERICAL RESULTS

In the following two examples, we attempt to demonstrate, wherever possible, that our bound works satisfactorily , which means that B^{**} in (5.3.18) and y_i^{**} in (5.3.7) do not differ much. The exact maximization in (5.3.7) was carried out numerically with one easy-to-use method of Gill and Murray (see [20]).

Example 5.4.1. Let

$$H_1 : Z(t) = .2\ Z(t-1) + .7\ Z(t-2) + \varepsilon(t)$$
$$H_2 : Z(t) = .5Z\ (t-1) + .3\ Z(t-2) + \varepsilon(t)$$

where $\{Z(t),\ t\geq o\}$ is as in Example 3.6.1.

The first row of R_1 is given by (for n = 10) :

(1, .666, .833, .633, .709, .585, .614, .532, .536, .480)

The first row of R_2 is given by (for n = 10) :

(1 , .714, .657, .543, .468, .397, .339, .289, .246, .209).

The computations are shown in Table 5.1. We note that

$$B^{**} = \max\left\{\ \frac{(\lambda_{max}(R_2^{-1}))^{1/2}}{(\lambda_{max}(R_2^{-1}R_1))^{1/2}+1}\ ,\ \frac{(\lambda_{max}(R_1^{-1}))^{1/2}}{(\lambda_{max}(R_1^{-1}R_2))^{1/2}+1}\ \right\} = \frac{(\lambda_{max}(R_2^{-1}))^{1/2}}{(\lambda_{max}(R_2^{-1}R_1))^{1/}}$$

Example 5.4.2. Let H_1 and H_2 be the same as in Example 3.6.2. In this case $B^{**} = (\lambda_{max}(R_1^{-1}))^{1/2}/((\lambda_{max}(R_1^{-1}R_2))^{1/2}+1)$. The results are shown in Table 5.2.

Table 5.1				Table 5.2		

Computations in Examples 5.4.1 and 5.4.2

n	B^{**}	y_1^{**}		n	B^{**}	y_1^{**}
2	.89870	.89870		2	0.68217	0.68217
3	1.01676	1.03571		3	0.72471	0.76165
4	1.02888	1.11896		4	0.77158	0.83835
5	1.12974	1.13848		5	0.81398	0.86866
6	1.13499	1.19251		6	0.85997	0.89182
7	1.15986	1.19353		7	0.86654	0.91101
8	1.20804	1.22125		8	0.87163	0.91201
9	1.22292	1.22492		9	0.89673	0.91986
10	1.23346	1.24716		10	0.90164	0.92227

5.5 CONCLUSION

We see that even in the simple case when the two types
of error are made equal, we do not obtain an explicit expression
for the optimum alpha vector and the optimum signal. An analytical
solution for the optimum signal is available only through a bound
on the total probability of misclassification. However, this
method seems to be reasonably good as a comparison with the
exact optimum value speaks for this method.

CHAPTER VI

LINEAR DISCRIMINANT FUNCTIONS AND DESIGN OF SIGNALS
FOR COMPLEX NORMAL TIME SERIES

6.1 INTRODUCTION

So far we have concentrated on the real normal time series. Since complex normal processes are of interest in many applied areas ([40]), we consider the linear discriminant functions for complex normal time series and try to extend all our major results from the real to the complex case. Though its origin lies in engineering and physical sciences, the complex stochastic processes, the complex normal process, in particular, have been extensively studied by statisticians ([58]). In Section 6.2 we formulate the problem of finding a classification rule based on linear statistics which maximizes the Bhattacharyya distance. Results analogous to those in the case of real discrete time series are obtained in Section 6.3 whereas Section 6.4 extends the same to a continuous time series. The problem of design of signals is considered in Section 6.5. Some basic definitions on complex normal processes are given in Chapter II.

6.2 FORMULATION OF THE PROBLEM

The problem is to classify an n-dimensional observation $z = (Z(o), \ldots, Z(n-1))$ as coming from one of the two categories

specified by two hypotheses H_1 and H_2. These hypotheses state that the nx1 complex normal time series z has the following means and covariances under H_1 and H_2 :

$$E_{H_j} z = \mu_j$$

and
$$E_{H_j} (z-\mu_j)(\bar{z}-\bar{\mu}_j)' = R_j \quad (j = 1,2)$$

The density of z is given by (under H_j),

$$p_j(z) = \frac{1}{\pi^n | R_j |} \exp\{ -(\bar{z} - \bar{\mu}_j)' R_j^{-1} (z - \mu_j)\}, \quad (j=1,2) \quad (6.2.1$$

We assume R_j to be Hermitian positive definite covariance matrices, and $\mu_1 \neq \mu_2$.

Let us look at the form of the likelihood ratio when $R_1 = R_2 (=R)$. We have,

$$\frac{p_1(z)}{p_2(z)} = \exp \{ \bar{z}' R^{-1} z - \bar{\mu}_2 R^{-1} z - \bar{z}' R^{-1} \mu_2 + \bar{\mu}_2' R^{-1} \mu_2 - \bar{z}' R^{-1} z$$
$$+ \bar{\mu}_1' R^{-1} z + \bar{z}' R^{-1} \mu_1 - \bar{\mu}_1' R^{-1} \mu_1 \}$$

$$= \exp \{ (\bar{\mu}_1' R^{-1} z - \bar{\mu}_2' R^{-1} z) + (\bar{z}' R^{-1} \mu_1 - \bar{z}' R^{-1} \mu_2)$$
$$+ (\bar{\mu}_2' R^{-1} \mu_2 - \mu_1' R^{-1} \bar{\mu}_1) \}$$

$$= \exp \left[\operatorname{Re}\{ (\bar{\mu}_1 - \bar{\mu}_2)' R^{-1} z \} + (\mu_2' R \bar{\mu}_2 - \mu_1' R^{-1} \bar{\mu}_1) \right]$$

In the light of this, coupled with the fact that the distributional theory associated with the quadratic discriminant function in the case of unequal covariances is very complicated ([40]), we consider the following linear procedures :

$$\text{Re}(\underset{\sim}{\alpha}^* \underset{\sim}{z}) \underset{\underset{H_2}{<}}{\overset{H_1}{\overset{>}{}}} c \qquad (6.2.\,)$$

$$\text{(accept)}$$

where $\underset{\sim}{\alpha}_{n \times 1}$ is an n-dimensional complex vector, and c is a scalar. We consider the solutions for $\underset{\sim}{\alpha}$ corresponding to the criterion of maximizing the Bhattacharyya distance.

6.3 DISCRITE TIME SERIES

6.3.1 METHOD OF FINDING THE ALPHA VECTOR FOR AN ARBITRARY TIME SERIES

First we note that $\underset{\sim}{\alpha}^* \underset{\sim}{z}$ is one dimensional complex random variable distributed normally with mean $\underset{\sim}{\alpha}^* \underset{\sim}{\mu}_j$ and variance $\underset{\sim}{\alpha}^* R_j \underset{\sim}{\alpha}$ under $H_j (j = 1,2)$ (see [40]). Hence (see Chapter II),

$$\zeta \overset{\Delta}{=} \text{Re}(\underset{\sim}{\alpha}^* \underset{\sim}{z}) \sim N(\text{Re}(\underset{\sim}{\alpha}^* \underset{\sim}{\mu}_j), \tfrac{1}{2} \underset{\sim}{\alpha}^* R_j \underset{\sim}{\alpha}) \text{ under } H_j(j=1,2)$$

Thus,

$$\rho_2(1,2;\zeta) = \frac{(\tfrac{1}{2} \underset{\sim}{\alpha}^* R_1 \underset{\sim}{\alpha} \cdot \tfrac{1}{2} \underset{\sim}{\alpha}^* R_2 \underset{\sim}{\alpha})^{1/4}}{\{\tfrac{1}{4} \underset{\sim}{\alpha}^*(R_1 + R_2)\underset{\sim}{\alpha}\}^{\frac{1}{2}}} \exp\{-\tfrac{1}{2} \frac{(\text{Re}(\underset{\sim}{\alpha}^* \underset{\sim}{\delta}))^2}{\underset{\sim}{\alpha}^*(R_1 + R_2)\underset{\sim}{\alpha}}\} \quad (6.3.1)$$

where $\underset{\sim}{\delta} = \underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2$.

Differentiating (Appendix H) $\ln\rho_2(1,2,\zeta)$ with respect to $\underset{\sim}{\alpha}$ and setting $\frac{\partial \ln\rho_2}{\partial \underset{\sim}{\alpha}} = 0$, we get

$$\left[\frac{1}{\underset{\sim}{\alpha}^* R_1 \underset{\sim}{\alpha}} - \frac{2}{\underset{\sim}{\alpha}^*(R_1 + R_2)\underset{\sim}{\alpha}} + \frac{2(\text{Re}(\underset{\sim}{\alpha}^* \underset{\sim}{\delta}))^2}{(\underset{\sim}{\alpha}^*(R_1 + R_2)\underset{\sim}{\alpha})^2}\right] R_1' \bar{\underset{\sim}{\alpha}}$$

$$+ \left[\frac{1}{\underset{\sim}{\alpha}^* R_2 \underset{\sim}{\alpha}} - \frac{2}{\underset{\sim}{\alpha}^*(R_1 + R_2)\underset{\sim}{\alpha}} + \frac{2(\text{Re}(\underset{\sim}{\alpha}^* \underset{\sim}{\delta}))^2}{(\underset{\sim}{\alpha}^*(R_1 + R_2)\underset{\sim}{\alpha})^2}\right] R_2' \bar{\underset{\sim}{\alpha}}$$

$$= \frac{2\,\mathrm{Re}(\underset{\sim}{\alpha}^{*}\underset{\sim}{\delta})}{\underset{\sim}{\alpha}^{*}(R_1+R_2)\underset{\sim}{\alpha}}\underset{\sim}{\delta}$$

Thus, it follows from the same argument as used in the real case that the value of $\underset{\sim}{\alpha}$ for which $-\ln\rho_2(1,2;\zeta)$ is a maximum is given by

$$\underset{\sim}{\alpha} = (R_1-\theta R_2)^{-1}\underset{\sim}{\delta}\,, \qquad\qquad (6.3.2)$$

where

$$-\theta = \frac{2\{\dfrac{\mathrm{Re}(\underset{\sim}{\alpha}^{*}\underset{\sim}{\delta})}{\underset{\sim}{\alpha}^{*}(R_1+R_2)\underset{\sim}{\alpha}}\}^2 + \dfrac{1}{\underset{\sim}{\alpha}^{*}R_2\underset{\sim}{\alpha}} - \dfrac{2}{\underset{\sim}{\alpha}^{*}(R_1+R_2)\underset{\sim}{\alpha}}}{2\{\dfrac{\mathrm{Re}(\underset{\sim}{\alpha}^{*}\underset{\sim}{\delta})}{\underset{\sim}{\alpha}^{*}(R_1+R_2)\underset{\sim}{\alpha}}\}^2 + \dfrac{1}{\underset{\sim}{\alpha}^{*}R_1\underset{\sim}{\alpha}} - \dfrac{2}{\underset{\sim}{\alpha}^{*}(R_1+R_2)\underset{\sim}{\alpha}}} \qquad (6.3.3)$$

An iterative procedure must be employed to solve for α.

Remark 6.3.1 : Since there always exists a non-singular matrix P such that

$$R_1 = P^{*}P$$

and

$$R_2 = P^{*}\Lambda P$$

where Λ is diagonal with elements as the eigen values of $R_2 R_1^{-1}$, a similar study on convergence of the interation involved in (6.3.2) can be carried out as we have done in the real case.

6.3.2 METHOD FOR FINDING $\underset{\sim}{\alpha}$ IN THE CASE OF LARGE SAMPLE FOR COVARIANCE STATIONARY TIME SERIES

The following theorem gives an explicit expression for the optimal vector $\underset{\sim}{\alpha}$ in the sense of maximizing the Bhattacharyya

distance asymptotically under certain regularity conditions similar to those made in the real case.

Theorem 6.3.1 : Let

(1) the n-dimensional vector $\underset{\sim}{z} = (z(o),\ldots,z(n-1))$ be a covariance stationary complex normal time series with mean $\underset{\sim}{\mu}_j$ and covariance matrix $R_j = ((r_j(s-t),s,t = 0,\ldots,n-1))$ under hypotheses H_j $(j = 1,2)$,

(2) the spectral densities $W_j(\lambda)$ of the process under the hypotheses are positive on $[-\pi,\pi]$,

(3) the sequence of mean difference $\delta(t)$ satisfies

(i) $\sup_t |\delta(t)| < \infty$

(ii) $\xi_n(\tau) \triangleq \frac{1}{n} \sum_{t=o}^{n-1-|\tau|} \delta(t) \overline{\delta(t+|\tau|)}$ has a limit

given by

$$\xi(\tau) \triangleq \lim_{n \to \infty} \xi_n(\tau) = \int_{-\pi}^{\pi} e^{i\lambda\tau} \frac{dU(\lambda)}{2\pi} \ ,$$

where $U(\lambda)$ plays the role of $M(\lambda)$ as described in the real case,

(4) $\sum_{t = -\infty}^{\infty} |t|^{1+\beta} |r_j(t)| < \infty$

where $o < \beta < 1$ $(j = 1,2)$.

Then the desired optimal $\underset{\sim}{\alpha}$ is given by

$$\underset{\sim}{\alpha} = (R_1+R_2)^{-1}\underset{\sim}{\delta}$$

Proof : We note that,

$$\ln\rho_2(1,2,\zeta_\theta) = \frac{1}{4} \ln \underset{\sim}{\delta}' R_\theta^{-1} R_1 R_\theta^{-1} \underset{\sim}{\delta} + \frac{1}{4} \ln \underset{\sim}{\delta}' R_\theta^{-1} R_2 R_\theta^{-1} \underset{\sim}{\delta}$$

$$- \frac{1}{2} \ln \underset{\sim}{\delta}' R_\theta^{-1} (P_1+P_2) R_\theta^{-1} \underset{\sim}{\delta} - \frac{1}{2} \frac{(\underset{\sim}{\delta}' R_\theta^{-1}\underset{\sim}{\delta})^2}{\underset{\sim}{\delta}' R_\theta^{-1} (R_1+R_2) R_\theta^{-1} \underset{\sim}{\delta}}$$

The rest follows if we use the same approach as in the real case.

The following remark is inevitable.

Remark 6.3.2 : The linear procedure (6.2.2) gives rise to two types of errors which are as follows :

$$e_1 = Pr_1(Re(\underset{\sim}{\alpha}^*\underset{\sim}{z}) < c) = 1 - \Phi(y_1), \quad \text{where} \quad y_1 \overset{\Delta}{=} \frac{Re(\underset{\sim}{\alpha}^*\underset{\sim}{\mu}_1) - c}{(\frac{1}{2}\underset{\sim}{\alpha}^*R_1\underset{\sim}{\alpha})^{\frac{1}{2}}}$$

and $e_2 = Pr_2(Re(\underset{\sim}{\alpha}^*\underset{\sim}{z}) > c) = 1 - \Phi(y_2)$, where $y_2 \overset{\Delta}{=} \dfrac{c - Re(\underset{\sim}{\alpha}^*\underset{\sim}{\mu}_2)}{(\frac{1}{2}\underset{\sim}{\alpha}^*R_2\underset{\sim}{\alpha})^{\frac{1}{2}}}$

Thus,

$$y_2 = \frac{Re(\underset{\sim}{\alpha}^*\underset{\sim}{\delta}) - y_1(\frac{1}{2}\underset{\sim}{\alpha}^*R_1\underset{\sim}{\alpha})^{\frac{1}{2}}}{(\frac{1}{2}\underset{\sim}{\alpha}^*R_2\underset{\sim}{\alpha})^{\frac{1}{2}}}$$

Suppose e_1 is given or equivalently y_1. Then one can show very easily that $y_2 \longrightarrow \infty$ as $n \longrightarrow \infty$ when $\underset{\sim}{\alpha} = (R_1 + R_2)^{-1}\underset{\sim}{\delta}$.

6.3.3 SOME SPECIAL CATEGORIES OF PROBLEMS

1) Suppose $\underset{\sim}{\delta}$ is a null vector. Then from (6.3.1)

$$\rho_2(1,2;\underset{\sim}{z}) = \frac{\{(\frac{1}{2}\underset{\sim}{\alpha}^*R_1\underset{\sim}{\alpha})(\frac{1}{2}\underset{\sim}{\alpha}^*R_2\underset{\sim}{\alpha})\}^{\frac{1}{4}}}{\{\frac{1}{4}\underset{\sim}{\alpha}^*(R_1 + R_2)\underset{\sim}{\alpha}\}^{\frac{1}{2}}} \qquad (6.3.4)$$

Since there always exists a non-singular matrix P such that

$$R_1 = P^* P$$
$$R_2 = P^* \Lambda P$$

where Λ is diagonal with elements as the eigen values of the matrix $R_2 R_1^{-1}$; we can rewrite (6.3.4) as

$$\rho_2(1,2,\zeta) = (2)^{\frac{1}{2}} \frac{(\underset{\sim}{\beta}^* \Lambda \underset{\sim}{\beta} / \underset{\sim}{\beta}^* \underset{\sim}{\beta})^{\frac{1}{4}}}{(1 + \dfrac{\underset{\sim}{\beta}^* \Lambda \underset{\sim}{\beta}}{\underset{\sim}{\beta}^* \underset{\sim}{\beta}})^{\frac{1}{2}}}$$

where $\qquad P\underset{\sim}{\alpha} = \underset{\sim}{\beta}$.   (6.3.5)

By a similar argument as in the real case, we have the following.

Theorem 6.3.2 : The optimal $\underset{\sim}{\beta}$ is the eigen vector corresponding to $\lambda_{min}(R_2 R_1^{-1})$ or $\lambda_{max}(R_2 R_1^{-1})$ according as

$$\lambda_{min}(R_2 R_1^{-1}) \cdot \lambda_{max}(R_2 R_1^{-1}) \lessgtr 1.$$

The optimal $\underset{\sim}{\alpha}$ is obtained by solving (6.3.5).

2) Let

$$Z(t) = \begin{cases} b\,\mu(t) + \tilde{n}(t) \text{ , under } H_1 \\[2em] \tilde{n}(t), \text{ under } H_2, \end{cases}$$

where b is a complex normal random variable with zero mean and $E|b|^2 = 1$, $\mu(t)$ is a complex-valued deterministic function of t, $\tilde{n}(t)$ is a zero mean complex normal process. The above model corresponds to the radar problem where we have to decide whether or not a target is present at a particular location ([64]).

Let n observations be made on $\{Z(t),\ t\epsilon T\}$ and

$$\underset{\sim}{z} = (Z(o),Z(1),\ldots,\ Z(n-1))$$

$$\underset{\sim}{\mu} = (\mu(o),\ldots\ldots\ldots,\ \mu(n-1))$$

$$\underset{\sim}{n} = (\widetilde{n}(o),\ldots\ldots\ldots,\ \widetilde{n}(n-1))$$

Then $E\underset{\sim}{z} = 0$, $R_1 = E_{H_1}\underset{\sim\sim}{zz}{}^* = E_{H_1}(b\underset{\sim}{\mu} + \underset{\sim}{n})(b\underset{\sim}{\mu} + \underset{\sim}{n})^*$

$$= R_2 + \underset{\sim\sim}{\mu\mu}{}^*,$$

where $\qquad E\ \underset{\sim\sim}{nn}{}^* = R_2.$

Then (6.3.1) reduces to

$$\rho_2(1,2,\zeta) = (2)^{\frac{1}{2}} \frac{\left[\ \{\underset{\sim}{\alpha}{}^*(R_2+\underset{\sim\sim}{\mu\mu}{}^*)\underset{\sim}{\alpha}\}\ \{\underset{\sim}{\alpha}{}^*(R_2)\underset{\sim}{\alpha}\}\ \right]^{\frac{1}{4}}}{\{\underset{\sim}{\alpha}{}^*(2R_2 + \underset{\sim\sim}{\mu\mu}{}^*)\underset{\sim}{\alpha}\}^{\frac{1}{2}}}$$

$$= \frac{(1 + \dfrac{|\underset{\sim}{\alpha}{}^*\underset{\sim}{\mu}|^2}{\underset{\sim}{\alpha}{}^*R_2\underset{\sim}{\alpha}})^{\frac{1}{4}}}{(1 + \dfrac{1}{2}\dfrac{|\underset{\sim}{\alpha}{}^*\underset{\sim}{\mu}|^2}{\underset{\sim}{\alpha}{}^*R_2\underset{\sim}{\alpha}})^{\frac{1}{2}}} \qquad (6.3.6)$$

Then we have the following

Theorem 6.3.3 : The optimal $\underset{\sim}{\alpha}$ in the sense of maximizing $-\ln\rho_2(1,2\zeta)$ is given by

$$\underset{\sim}{\alpha}{}_* = R^{-1}\ \underset{\sim}{\mu} \qquad (6.3.7)$$

Proof : Noting (6.3.6) is of the form

$$y = \frac{(1+x)^{\frac{1}{4}}}{(1+\frac{1}{2}x)^{\frac{1}{2}}}$$

where y decreases as x increases, by virtue of the following inequality

$$|\underset{\sim}{\alpha}{}^{*}\underset{\sim}{\mu}|^{2} \leq (\underset{\sim}{\alpha}{}^{*}R_{2}\underset{\sim}{\alpha})(\underset{\sim}{\mu}{}^{*}R_{2}^{-1}\underset{\sim}{\mu}),$$

(see ([57]), the theorem follows at once.

3) Let $\underset{\sim}{z} = \begin{cases} \underset{\sim}{\mu} + \underset{\sim}{n} , \text{ under } H_1 \\ \underset{\sim}{n} , \text{ under } H_2 \end{cases}$

where $\underset{\sim}{\mu}$ has a n-variate complex (real) normal distribution with zero mean and covariance matrix R_3, and $\underset{\sim}{n}$ has a n-variate normal distribution with zero mean and covariance matrix R_2, independent of $\underset{\sim}{\mu}$.

Then $R_1 \overset{\Delta}{=} E_{H_1} \underset{\sim}{z} \underset{\sim}{z}{}^{*} = R_2 + R_3$.
Consequently, (6.3.4) reduces to

$$\rho_2(1,2;\zeta) = \frac{\{\frac{1}{2} \underset{\sim}{\alpha}{}^{*}(R_2+R_3)\underset{\sim}{\alpha}\}^{\frac{1}{4}} \{\frac{1}{2} \underset{\sim}{\alpha}{}^{*}R_2\underset{\sim}{\alpha}\}^{\frac{1}{4}}}{\{\frac{1}{4} \underset{\sim}{\alpha}{}^{*}(2R_2+R_3)\underset{\sim}{\alpha}\}^{\frac{1}{2}}}$$

$$= (1 + \frac{\underset{\sim}{\alpha}{}^{*}R_3\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}{}^{*}R_2\underset{\sim}{\alpha}})^{\frac{1}{4}}/(1 + \frac{1}{2} \frac{\underset{\sim}{\alpha}{}^{*}R_3\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}{}^{*}R_2\underset{\sim}{\alpha}})^{\frac{1}{2}} \qquad (6.3.8)$$

Then we have the following theorem.

Theorem 6.3.4 : The optimum $\underset{\sim}{\alpha}$ in the sense of maximizing the $-\ln\rho_2(1,2;\chi)$ is the eigen vector corresponding to the maximum eigen value of $R_2^{-1}R_3$.

Proof : It follows once we observe that (6.3.8) is of the form

$$y = \frac{(1+x)^{\frac{1}{4}}}{(1+\frac{1}{2}x)^{\frac{1}{2}}}$$

where $x > 0$.

6.4   CONTINUOUS TIME SERIES

6.4.1 SECOND ORDER TIME SERIES

The observed time series $\{Z(t), t\varepsilon T\}$ is assumed to be continuous in time. Let $T = [o,A]$, where A is a real number (finite). Assume $E|Z(t)|^2 < \infty$ for all $t\varepsilon T$. Let

$$E_{H_j} Z(t) = \mu_j(t) \tag{6.4.1}$$

$$\text{Cov}_{H_j}(Z(t), \overline{Z(u)}) = R_j(t,u) \tag{6.4.2}$$

where $(j = 1,2)$ and $t,u\varepsilon T$.

Our first step is to reduce the process $\{Z(t), t\varepsilon T\}$ to a set of random variables (possibly countably infinite set). This is achieved by the method of the series expansion :

$$Z(t) = \sum_{n=1}^{\infty} z_n \varphi_n(t) \tag{6.4.3}$$

in the following sense :

$$\lim_{K \to \infty} E|Z(t) - \sum_{n=1}^{K} z_n \varphi_n(t)|^2 = o, \quad 0 \le t \le A$$

where $z_n = \int_0^A Z(t)\varphi_n(t)dt$ (6.4.4)

and $\{\varphi_n(t)\}$ is a set of complete complex orthonormal functions in the interval $[o,A]$ i.e.

$$\int_0^A \varphi_n(t) \overline{\varphi_m(t)}dt = \delta_{nm} ,$$

(see Appendix F).

Following the arguments as in the real case, the linear procedures of interest are given by

$$\zeta \stackrel{\Delta}{=} Re(\int_0^A \alpha^*(t)Z(t)dt) \underset{H_2}{\overset{H_1}{\gtrless}} c$$ (6.4.5)
$$(\text{accept})$$

(see Appendix I for a definition of stochastic integral of a complex process). The problem is now to find an optimal $\alpha(t)$, $t\varepsilon[o,A]$ in the sense of maximizing $-\ln\rho_2(1,2;\zeta)$, where ζ is defined in (6.4.5).

The following theorem states the method for obtaining $\alpha(t),t\varepsilon[o,A]$

<u>Theorem 6.4.1</u> : The $\alpha(t)$, $t\varepsilon[o,A]$, for which $-\ln\rho_2(1,2;\zeta)$ is a maximum satisfies an integral equation of the following type,

$$\int_0^A R_\theta(t,u)\alpha(t)dt = \delta(u),$$ (6.4.6)

to be solved iteratively, where θ is given by

$$-\theta = \cfrac{\left\{\cfrac{\text{Re}(\int_0^A \alpha^*(t)\delta(t)dt)}{\int_0^A\int_0^A R(t,u)\alpha^*(t)\alpha(u)dtdu}\right\}^2 + \cfrac{1}{\int_0^A\int_0^A R_2(t,u)\alpha^*(t)\alpha(u)dtdu} - \cfrac{2}{\int_0^A\int_0^A R(t,u)\alpha^*(t)\alpha(u)dtdu}}{\left\{\cfrac{\text{Re}(\int_0^A \alpha^*(t)\delta(t)dt)}{\int_0^A\int_0^A R(t,u)\alpha^*(t)\alpha(u)dtdu}\right\}^2 + \cfrac{1}{\int_0^A\int_0^A R_1(t,u)\alpha^*(t)\alpha(u)dtdu} - \cfrac{2}{\int_0^A\int_0^A R(t,u)\alpha^*(t)\alpha(u)dtdu}}$$

and
$$R_\rho(t,u) \overset{\Delta}{=} R_1(t,u) - \theta R_2(t,u)$$
$$R(t,u) \overset{\Delta}{=} R_1(t,u) + R_2(t,u)$$

The iteration continues until $|\theta^{(i+1)}-\theta^{(i)}| < \varepsilon$, where ε is a pre-assigned number and i denotes the number of the iteration.

Proof : We adopt the approach taken in proving **Theorem 4.2.1.**

6.4.2 COVARIANCE STATIONARY TIME SERIES WITH THE INDEX SET [o,∞)

The following theorem gives an explicit expression for $\alpha(t)$, $0 \le t<\infty$, if $Z(t)$ is a stationary time series.

Theorem 6.4.2 : Let

$$S_j(\lambda) \overset{\Delta}{=} \int_0^\infty V_j(v)\, e^{i\lambda v}\, dv \qquad (6.4.7)$$

$$D_j(\lambda) \overset{\Delta}{=} \int_0^\infty \delta_j(u)\, e^{i\lambda u}\, du, \quad (j = 1,2) \qquad (6.4.8)$$

where
$$V_1(v) \overset{\Delta}{=} \text{Re}\, R_1(v) + \text{Re}\, R_2(v) \qquad (6.4.9)$$

$$V_2(v) \overset{\Delta}{=} \text{Im } R_1(v) + \text{Im } R_2(v) \tag{6.4.10}$$

and
$$\delta(v) \overset{\Delta}{=} \delta_1(v) + i\delta_2(v) \tag{6.4.11}$$

Then under assumptions similar to those given in Theorem 4.3.1 with the obvious modifications required for the complex case, the optimal $\alpha(t)$ is given by

$$\alpha(t) = \int_{-\infty}^{\infty} (A_1(\lambda) + iA_2(\lambda)) e^{-i\lambda t} \frac{d\lambda}{2\pi} \tag{6.4.12}$$

where
$$A_1(\lambda) = \frac{S_1(\lambda)D_1(\lambda) + S_2(\lambda)D_2(\lambda)}{S_1^2(\lambda) + S_2^2(\lambda)} \tag{6.4.13}$$

$$A_2(\lambda) = \frac{S_2(\lambda)D_1(\lambda) - S_1(\lambda)D_2(\lambda)}{S_1^2(\lambda) + S_2^2(\lambda)} \quad , \quad (-\infty < \lambda < \infty)$$

$$\tag{6.4.14}$$

The following lemma which is known as the Sampling theorem for complex stochastic processes is essential in proving the above theorem.

Lemma 6.4.2 : If for some Λ (real number), the spectral density of a covariance stationary complex process $\{Z(t), 0 < t < \infty\}$ is zero outside of the frequency interval $-\Lambda \le \lambda \le \Lambda$, then

$$Z(t) = \sum_{k=0}^{\infty} \frac{\sin \Lambda (t - \frac{\Lambda k}{\Lambda})}{\Lambda (t - \frac{\pi k}{\Lambda})} Z(\frac{\pi k}{\Lambda}) \quad , \quad (0 < t < \infty)$$

We omit the proof as it is readily available in ([29],[7]).

<u>Proof of the Theorem 6.4.2</u> :

Using the approach as in Theorem 4.3.1, if we write

$$\alpha(t) = \alpha_1(t) + i\alpha_2(t),$$

$$\delta(t) = \delta_1(t) + i\delta_2(t),$$

$$R_1(t) = \text{Re}R_j(t) + i\text{Im}R_j(t), \ (j = 1,2),$$

we end up with the following simultaneous integral equations
after equating the real and imaginary parts of the equation
(6.4.6) :

$$\int_0^\infty V_1(t-u)\alpha_1(t)dt + \int_0^\infty V_2(t-u)\alpha_2(t)dt = \delta_1(u)$$

$$\int_0^\infty V_2(t-u)\alpha_1(t)dt - \int_0^\infty V_1(t-u)\alpha_2(t)dt = \delta_2(u)$$
(6.4.15)

By an application of Lemma 4.3.1 of Chapter IV, (6.4.15) reduces
to

$$S_1(\lambda)A_1(\lambda) + S_2(\lambda)A_2(\lambda) = D_1(\lambda)$$

$$S_2(\lambda)A_1(\lambda) - S_1(\lambda)A_2(\lambda) = D_2(\lambda)$$
(6.4.16)

From (6.4.16), the theorem follows at once.

6.4.3 SOME SPECIAL CATEGORIES OF PROBLEMS

1) Let

$$Z(t) = \begin{cases} b\mu(t) + \tilde{n}(t), & \text{under } H_1 \\ \tilde{n}(t), & \text{under } H_2 \end{cases}, \ 0 \le t \le A$$
(6.4.17)

where b, $\mu(t)$, $\tilde{n}(t)$ are as described in Section 6.3.3.
Then the following theorem states how to obtain the optimum $\alpha(t)$ in
the sense of maximizing $-\ln P_2(1,2;\zeta)$.

<u>Theorem 6.4.3</u> : The optimum $\alpha(t)$, $0 \leq t \leq A$, can be obtained by solving the following integral equation :

$$\int_0^A R_2(t,u)\alpha(t)dt = \mu(u) \tag{6.4.18}$$

<u>Proof</u> : We adopt the approach taken in proving Theorem 4.2.1.

2) Let $\tilde{n}(t)$ contain two statistically independent components $\tilde{n}_c(t)$ and $\tilde{\omega}(t)$, where $\tilde{\omega}(t)$ is a white complex normal process with

$$E \tilde{\omega}(t) \tilde{\omega}^*(u) = \delta_D(t-u).$$

Then the model (6.4.17) can be recast in the form :

$$Z(t) = \begin{cases} b\mu(t) + \tilde{n}_c(t) + \tilde{\omega}(t) \\ \\ \tilde{n}_c(t) + \tilde{\omega}(t) \end{cases} \tag{6.4.19}$$

We consider a special form of this model in which

$$\tilde{n}_c(t) = \sum_{j=1}^{K} \tilde{b}_j \, \mu(t-\tau_j) \, e^{i\omega_j t}$$

Then (6.4.19) reduces to

$$Z(t) = \begin{cases} \tilde{b}_d\mu(t) + \sum\limits_{j=1}^{K} \tilde{b}_j\mu(t-\tau_j)e^{i\omega_j t} + \tilde{\omega}(t), \text{ under } H_1 \\ \\ \sum\limits_{j=1}^{K} \tilde{b}_j\mu(t-\tau_j)e^{i\omega_j t} + \tilde{\omega}(t), \text{ under } H_2, \ -\infty < t < \infty \end{cases} \tag{6.4.20}$$

where τ_j's, ω_j's and K are given constants, $\tilde{b}_d$ and $\tilde{b}_j$ are zero-mean complex normal r.v.s. that are statistically independent with unequal variances :

$$E \tilde{b}_d \tilde{b}_d^* = 2\sigma_d^2 \tag{6.4.21}$$

$$E \tilde{b}_j \tilde{b}_\ell^* = 2\sigma_j^2\delta_{j\ell} \quad , \quad j,\ell = 1,\ldots,K \tag{6.4.22}$$

and $\quad E\, \tilde{b}_d\, \hat{b}_d = E\, \tilde{b}_i \tilde{b}_i = E\, \tilde{b}_d \tilde{b}_j^* = E\, \tilde{b}_d \tilde{b}_j = 0 \ \forall\ j = 1,..,K$

$$(6.4.23)$$

The above model (6.4.20) corresponds to the problem (known as " Resolution problem in discrete environment " ([64])) of detecting a desired target in the presence of interfering targets. The interesting feature of this model is that the noise process depends an $\mu(t)$.

The covariance function of $\tilde{n}_c(t)$ is given by

$$\tilde{K}_c(t,u) = 2 \sum_{j=1}^{K} \sigma_j^2 \mu(t-\tau_j)\mu^*(u-\tau_j)e^{i\omega_j(t-u)} \qquad (6.4.24)$$

Since $R_2(t,u) = \tilde{K}_c(t,u) + \delta_D(t-u)$, using (6.4.24), (6.4.18) reduces to

$$
\begin{aligned}
\mu(t) &= \int_{-\infty}^{\infty} K_c(t,u)\alpha(u)du + \alpha(t) \\
&= \int_{-\infty}^{\infty} \{ \sum_{j=1}^{K} \mu(t-\tau_j)\mu^*(u-\tau_j)e^{i\omega_j(t-u)} \}\alpha(u)du + \alpha(t),
\end{aligned}
$$

$$(6.4.25)$$

$$= 2 \sum_{j=1}^{K} \sigma_j^2 \mu(t-\tau_j)\ e^{i\omega_j t}(\int_{-\infty}^{\infty} \mu^*(u-\tau_j)e^{-i\omega_j u}\alpha(u)du) + \alpha(t)$$

$$(6.4.26)$$

We note that (6.4.25) is an integral equation with a seperable kernel. The solution to (6.4.26) is given by

$$\alpha(t) = \mu(t) + \sum_{j=1}^{K} \alpha_j \mu(t-\tau_j)e^{i\omega_j t} \qquad (6.4.27)$$

where α_j's are constants to be determined. In what follows we describe a method for calculating the constants α_j (see [64]).

Define
$$\underset{\sim}{b} \overset{\Delta}{=} (\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_K)'$$

$$\underset{\sim}{\mu}_\tau(t) \overset{\Delta}{=} (\mu(t-\tau_1)e^{i\omega_1 t}, \ldots, \mu(t-\tau_K)e^{i\omega_K t})'$$

$$\tilde{\Lambda} = E \underset{\sim}{b} \underset{\sim}{b}^* = 2 \begin{bmatrix} \sigma_1^2 & & O \\ & \ddots & \\ O & & \sigma_K^2 \end{bmatrix}$$

and
$$\tilde{\rho} \overset{\Delta}{=} \int_{-\infty}^{\infty} \underset{\sim}{\mu}_\tau(t) \, \underset{\sim}{\mu}_\tau^*(t) dt$$

Thus (6.4.24) can be rewritten as

$$\tilde{K}_c(t,u) = \underset{\sim}{\mu}_\tau'(t) \tilde{\Lambda} \underset{\sim}{\mu}_\tau^*(u) \tag{6.4.28}$$

We can write (6.4.27) in the matrix notation as

$$\alpha(t) = \mu(t) + \underset{\sim}{\alpha}' \underset{\sim}{\mu}_\tau(t)$$
$$= \mu(t) + \underset{\sim}{\mu}_\tau'(t)\underset{\sim}{\alpha} , \tag{6.4.29}$$

where
$$\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)'.$$

Substituting (6.4.28) and (6.4.29) in (6.4.25) , we have

$$\mu(t) = \int_{-\infty}^{\infty} \{ \underset{\sim}{\mu}_\tau'(t)\tilde{\Lambda} \underset{\sim}{\mu}_\tau^*(u) + \delta_D(t-u)\} \{ \underset{\sim}{\mu}(u) + \underset{\sim}{\mu}_\tau'(u)\underset{\sim}{\alpha} \} du$$

$$= \underset{\sim}{\mu}_\tau'(t) \tilde{\Lambda} \tilde{\rho}_d + \mu(t) + \underset{\sim}{\mu}_\tau'(t) \tilde{\Lambda} \tilde{\rho}^* \underset{\sim}{\alpha} + \underset{\sim}{\mu}_\tau'(t)\underset{\sim}{\alpha} , \tag{6.4.30}$$

where
$$\tilde{\rho}_d \overset{\Delta}{=} \int_{-\infty}^{\infty} \underset{\sim}{\mu}_\tau^*(u)\mu(u) du \tag{6.4.31}$$

The required $\underset{\sim}{\alpha}$ is thus given by

$$\underset{\sim}{\alpha} = -(I + \tilde{\Lambda} \tilde{\rho}^*)^{-1} \tilde{\Lambda} \tilde{\rho}_d \tag{6.4.32}$$

The above result is illustrated by a simple example in Van Trees ([64]).

<u>Remark 6.4.1</u> : Let us consider the model as in (6.4.17), Then the $\alpha(t)$ which maximizes $-\ln\rho_2(1,2;\zeta)$ satisfies (6.4.18), that is,

$$\int_0^A R_2(t,u)\ \alpha(t)\ dt = \mu(u) \tag{6.4.33}$$

Now, our test is

$$\zeta = Re(\ \int_0^A \alpha^*(t)Z(t)dt) \underset{H_2}{\overset{H_1}{\gtrless}} c \tag{6.4.34}$$

The Bayes' optimum test is given by (see [64]),

$$\mid \int_0^A \tilde{g}^*(t)Z(t)dt \mid^2 \underset{H_2}{\overset{H_1}{\gtrless}} \gamma$$

where

$$\gamma \overset{\Delta}{=} (\ln m - \ln \frac{\omega}{1-\omega})\ \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2-\sigma_2^2}$$

and $\tilde{g}$ satisfies

$$\int_0^A R_2(t,u)\tilde{g}(t)dt = \mu(u) \tag{6.4.35}$$

which is the same as (6.4.33).

6.5 DESIGN OF SIGNALS

1) Model I

Let
$$Z(t) = \begin{cases} \mu(t) + n_1(t)\ ,\ \text{under } H_1 \\ n_2(t)\ ,\ \text{under } H_2\ ,\ t\varepsilon T \end{cases}$$

where $n_1(t)$ and $n_2(t)$ are two zero mean complex normal processes.

Set $e_1 = e_2$

$<=> \ Y_1 = Y_2$

$$=> \quad Y_1 = \frac{Re(\underset{\sim}{\alpha}^* \underset{\sim}{\mu})}{(\underset{\sim}{\alpha}^* R_1 \underset{\sim}{\alpha})^{\frac{1}{2}} + (\underset{\sim}{\alpha}^* R_2 \underset{\sim}{\alpha})^{\frac{1}{2}}}$$

Then we have the following theorem analogous to that in the real case.

<u>Theorem 6.5.1</u> : Our signal selection criterion is

$$\underset{\substack{\underset{\sim}{\mu} \\ \underset{\sim}{\mu}^* \underset{\sim}{\mu} = 1}}{max} \quad \underset{\underset{\sim}{\alpha}}{max} \quad \frac{\underset{\sim}{\alpha}^* \underset{\sim}{\mu}}{(\underset{\sim}{\alpha}^* R_j \underset{\sim}{\alpha})^{\frac{1}{2}}} \quad (j = 1, \text{ or } 2)$$

The optimal $\underset{\sim}{\alpha}$ and $\underset{\sim}{\mu}$ are given by

$$\underset{\sim}{\alpha}_* = R_1^{-1} \underset{\sim}{\mu}_* \quad \text{or} \quad R_2^{-1} \underset{\sim}{\mu}_*$$

and $\underset{\sim}{\mu}_*$ is an eigen-vector corresponding to the $\lambda_{min}(R_1)$ or $\lambda_{min}(R_2)$ according as

$$\frac{(\lambda_{max}(R_1^{-1}))^{\frac{1}{2}}}{(\lambda_{max}(R_1^{-1} R_2))^{\frac{1}{2}} + 1} \gtrless \frac{(\lambda_{max}(R_2^{-1}))^{\frac{1}{2}}}{(\lambda_{max}(R_2^{-1} R_1))^{\frac{1}{2}} + 1}$$

<u>Proof</u> : It follows immediately if we note that the results of ([57]) used in proving Theorem 5.3.1 hold good in the complex case also.

<u>Remark 6.5.1</u> : If we set $e_1 = k_1 e_2$, the same result follows.

Model II :

The following model occurs in communication.

$$Z(t)= \begin{cases} (b+k)\mu(t) + \tilde{n}(t) \text{ , under } H_1 \\ \\ \tilde{n}(t) \text{ , under } H_2, \end{cases}$$

where k is a given constant, b, $\mu(t)$, $\tilde{n}(t)$ are the same as in (6.4.17). The following theorem states what is the optimal signal

__Theorem 6.5.2__ : Our criterion is : $\min_{e_1=e_2} Pr(\varepsilon)$. Then the optimal signal is the eigen vector corresponding to $\lambda_{min}(R_2)$.

__Proof__ : Set $y_1 = y_2$

$$\Rightarrow y_1 = \frac{k \ Re(\underset{\sim}{\alpha}^*\underset{\sim}{\mu})}{(\frac{1}{2} \underset{\sim}{\alpha}^* R_1 \underset{\sim}{\alpha})^{\frac{1}{2}} + (\frac{1}{2} \underset{\sim}{\alpha}^* R_2 \underset{\sim}{\alpha})^{\frac{1}{2}}}$$

$$= \frac{(2)^{\frac{1}{2}} k \ Re(\underset{\sim}{\alpha}^* \underset{\sim}{\mu})}{(\underset{\sim}{\alpha}^*(R_2 + k^2 \underset{\sim}{\mu}\underset{\sim}{\mu}^*)\underset{\sim}{\alpha})^{\frac{1}{2}} + (\underset{\sim}{\alpha}^* R_2 \underset{\sim}{\alpha})^{\frac{1}{2}}}$$

$$= \frac{(2)^{\frac{1}{2}} k \ Re(\underset{\sim}{\alpha}^* \underset{\sim}{\mu})/(\underset{\sim}{\alpha}^* R_2 \underset{\sim}{\alpha})^{\frac{1}{2}}}{1 + (1 + \frac{k^2|\underset{\sim}{\alpha}^* \underset{\sim}{\mu}|^2}{\underset{\sim}{\alpha}^* R_2 \underset{\sim}{\alpha}})^{\frac{1}{2}}} \qquad (6.5.1)$$

Since the optimal $\underset{\sim}{\alpha}$ is of the form :

$$\underset{\sim}{\alpha} = R_\theta^{-1} \underset{\sim}{\delta} \ ,$$

without loss of generality we can take $\underset{\sim}{\alpha}^{*}\underset{\sim}{\mu}$ to be real.

Thus, (6.5.1) reduces to

$$Y_1 = \frac{(2)^{\frac{1}{2}} k(\underset{\sim}{\alpha}^{*}\underset{\sim}{\mu})/(\underset{\sim}{\alpha}^{*}R_2\underset{\sim}{\alpha})^{\frac{1}{2}}}{1 + (1 + \frac{k^2 |\underset{\sim}{\alpha}^{*}\underset{\sim}{\mu}|^2}{\underset{\sim}{\alpha}^{*}R_2\underset{\sim}{\alpha}})^{\frac{1}{2}}}$$

which is of the form :
$$y = \frac{kx}{1 + (1+k^2x^2)^{\frac{1}{2}}}$$

where y increases as x increases. Thus the problem reduces to :

$$\underset{\substack{\underset{\sim}{\mu} \\ \underset{\sim}{\mu}^{*}\underset{\sim}{\mu}=1}}{\max} \quad \underset{\underset{\sim}{\alpha}}{\max} \quad \frac{\underset{\sim}{\alpha}^{*}\underset{\sim}{\mu}}{(\underset{\sim}{\alpha}^{*}R_2\underset{\sim}{\alpha})^{\frac{1}{2}}} ,$$

hence the theorem.

Remark 6.5.2 : The same result is obtained if our criterion is :
$$\underset{e_1 = k_1 e_2}{\min} \quad Pr(\varepsilon).$$

Remark 6.5.3 : The above results can easily be extended to the continuous case.

6.6 CONCLUSION

As we have seen in the real case, the compact form of the optimal linear discriminant function with respect to the criterion of maximizing the Bhattacharyya distance is available only in the asymptotic case if we consider the covariance stationary discrete time series and in the case when the continuous time series is

covariance stationary with the index set infinite. In some
other cases also explicit analytical forms of the LDFs are
available. In the case when the noise covariance kernel
is seperable finding the desired LDF reduces to solving
an algebraic system of equations.

Classification problem

QDF — I(1,2) — I(2,1) — J(1,2) — LDF: $a'z \geq z_c$

The proposed criterion: $-\ln p_2$

The Anderson procedure

Discrete time series

Stationary

Iterative scheme for finding optimal a; $a = R_2^{-1}\delta$

Convergence of the iteration process

One procedure is admissible

Comparison with other LDFs

Comparison with QDF

Solution in compact form for optimal a asymptotically

$\theta \xrightarrow{n\to\infty} -1$

$e_1, e_2 \xrightarrow{n\to\infty} 0$

Special categories of problems

Equivalence of the two classes of tests

Illustrative examples

Iterative scheme for finding $a(t)$, $t \in [0,\Lambda]$, Λ finite, via solving Fredholm type integral equations

Iterative scheme, $t \in [0,\Lambda]$, Converges

Special categories of problems

Stationary case: compact form for 3 asymptotic $t, s \ldots M \ldots$

$e_1, e_2 \to 0$ $n \to \infty$

Complex time series

Solution in compact form for $a(t)$, $t \in [0,\infty)$

Stationary

Discrete

Combinations

Design of Signals

Solution through a bound on the total probability of misclassification subject to the restriction $a_1 = a_2$ or more generally $a_1 = \kappa a_2$ (both real and complex case)

Some special categories of problems (complex)

Iterative scheme, $t \in [0,\Lambda]$

Special categories of problems

Stationary case: compact form for $a(t)$, $t \in [0,\infty)$

CHAPTER VII

CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

On page 193 we have given a Flow Chart of the work done in this thesis. We see in Chapter III that the proposed criterion of maximizing the Bhattacharyya distance for optimal LDF leads to solving an implicit equation for a discrete time series. The comparison of the performance of our LDF with other LDFs and QDF reveals that the test statistic of our interest is worth-considering. In the case of stationary time series, we observe that when no compact form of optimal LDFs is available for the criteria considered so far in the literature, the maximization of the Bhattacharyya distance does yield one asymptotically.

To the best of our knowledge, LDFs for continuous time series have not been considered so far in the literature. This issue is taken up in Chapter IV where we notice that finding the optimal LDF amounts to solving an integral equation of Fredholm type and we are able to obtain a compact form of the optimal LDF in the case when the time series is stationary with the observation interval infinite.

In ''Design of Signals'' problem considered in Chapter V, an analytical solution for the optimal signal is available only through a bound on the total probability of misclassificatio

In Chapter VI we are able to extend all the major results from the real to the complex-valued time series.

A wide variety of interesting problems arise out of the investigations carried out in this thesis. Some of them are mentioned below.

We have seen in Chapter III that an iterative scheme is employed to find a desired $\underset{\sim}{\alpha}$. It would be interesting to develop a recurrsive scheme for finding the same.

A good deal of effort is needed to make comparisons of the performance of our procedure with other LDFs as well as with QDF for continuous time series and to study the convergence of the iteration process given a Chapter IV.

We see in Chapter VI that under the model in (6.4.17) the integral equations yielding the desired $\alpha(t)$ of the linear procedure and the $\widetilde{g}(t)$ of the Bayes optimal test are identical. Further investigation is needed to determine the class of covariance matrices for which the above assertion holds good.

We have dealt with the two-group classification problem in the previous chapters. The classification problem involving more than two groups have to be looked into ; an approach adopted in [60] may be useful.

Suppose we have observations x_{ij} , $j = 1,2,..,n_i$, from H_i $(i = 1,2)$. Problem is to assign $\underset{\sim}{x}$ to H_1 or H_2 and $p(\underset{\sim}{x})$ is not known. The following approach is given in [58]. Calculate the distance between the sample distribution functions

of H_i ($i = 1,2$) for the two cases, the first including $\underset{\sim}{x}$ in H_1 and the second including $\underset{\sim}{x}$ in H_2. Assign $\underset{\sim}{x}$ to the hypothesis that gives the smaller distance. Effectiveness of the Bhattacharyya distance in this direction remains unexplored.

In the present work, we have considered normal processes. The area where the assumption of normality is no longer valid but the process of concern is best modeled by a Markov process (for example, AR processes) calls for further investigation ([63]).

APPENDIX   A

Bhattacharyya distance and the triangle inequality

In the following example (see [26]) we shall see that the Bhattacharyya distance need not satisfy the triangle inequality.

Take   three normal distributions with means zero and standard derivations 1,4, and 5, denoted by P_1, P_2, and P_3 respectively. Then we shall show :

$$-\ln\rho_2(P_1,P_2)-\ln\rho_2(P_2,P_3) < -\ln\rho_2(P_1,P_3) \qquad (1)$$

Now, $-\ln\rho_2(P_1,P_2) = \frac{1}{2} \ln \frac{1}{2}(1+16) - \frac{1}{4} \ln 16$

$$= 0.3769$$

$$-\ln\rho_2(P_2,P_3) = 0.0124$$

Thus,

$$-\ln\rho_2(P_1,P_2)- \ln\rho_2(P_2,P_3) = 0.3893$$

whereas $-\ln\rho_2(P_1,P_3) = 0.4778.$

From this, (1) follows.

APPENDIX B

Combining two Quadratic forms

Lemma

If $Q_j = (\underset{\sim}{x} - \underset{\sim}{\mu}_j)' R_j^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_j) (j = 1, 2)$, then

$$Q_1 + Q_2 = (\underset{\sim}{x} - \underset{\sim}{m})' S (\underset{\sim}{x} - \underset{\sim}{m}) + (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' (R_1 + R_2)^{-1} (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2) \qquad (1)$$

where

$$\underset{\sim}{m} = (R_1 + R_2)^{-1} (R_2 \underset{\sim}{\mu}_1 + R_1 \underset{\sim}{\mu}_2) \qquad (2)$$

and

$$S = R_1^{-1} (R_1 + R_2) R_2^{-1} \qquad (3)$$

Proof : Let $R_j^{-1} = A_j$, $(j = 1, 2)$ and $A = A_1 + A_2$.

Thus,

$$Q_1 + Q_2 = \underset{\sim}{x}' A \underset{\sim}{x} - 2 \underset{\sim}{x}' (A_1 \underset{\sim}{\mu}_1 + A_2 \underset{\sim}{\mu}_2) + \underset{\sim}{\mu}_1' A_1 \underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2' A_2 \underset{\sim}{\mu}_2$$

$$= \underset{\sim}{x}' A \underset{\sim}{x} - 2 \underset{\sim}{x}' A \underset{\sim}{m} + \underset{\sim}{m}' A \underset{\sim}{m} + \underset{\sim}{\mu}_1' A_1 \underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2' A_2 \underset{\sim}{\mu}_2 - \underset{\sim}{m}' A \underset{\sim}{m}$$

$$= (\underset{\sim}{x} - \underset{\sim}{m})' A (\underset{\sim}{x} - \underset{\sim}{m}) + \underset{\sim}{\mu}_1' A_1 \underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2' A_2 \underset{\sim}{\mu}_2 - \underset{\sim}{m}' A \underset{\sim}{m} \qquad (4)$$

where $\underset{\sim}{m}$ is defined in (2).

But $\underset{\sim}{m}' A \underset{\sim}{m} = (A_1 \underset{\sim}{\mu}_1 + A_2 \underset{\sim}{\mu}_2)' A^{-1} A A^{-1} (A_1 \underset{\sim}{\mu}_1 + A_2 \underset{\sim}{\mu}_2)$

$$= (A_1 \underset{\sim}{\mu}_1 + A_2 \underset{\sim}{\mu}_2)' A^{-1} (A_1 \underset{\sim}{\mu}_1 + A_2 \underset{\sim}{\mu}_2)$$

$$= [A_1 (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2) + A \underset{\sim}{\mu}_2]' A^{-1} [A \underset{\sim}{\mu}_1 - A_2 (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)]$$

$$= (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' A_1 A^{-1} A \underset{\sim}{\mu}_1 - (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' A_1 A^{-1} A_2 (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)$$

$$\quad + \underset{\sim}{\mu}_2' A A^{-1} A \underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2' A A^{-1} A_2 (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)$$

$$= (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' A_1 \underset{\sim}{\mu}_1 - (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' A_1 A^{-1} A_2 (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2) + \underset{\sim}{\mu}_2' A \underset{\sim}{\mu}_1$$

$$\quad - \underset{\sim}{\mu}_2' A_2 (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)$$

Thus,

$$\underset{\sim}{m}'A\underset{\sim}{m} = \mu_1'A_1\mu_1 + \mu_2'A_2\mu_2 - (\mu_1-\mu_2)'A_1A^{-1}A_2(\mu_1-\mu_2) \qquad (5)$$

Again,

$$A = A_1 + A_2 = R_1^{-1} + R_2^{-1} = R_1^{-1}(R_1+R_2)R_2^{-1} = R_2^{-1}(R_2+R_1)R_1^{-1}$$

$$(6)$$

and $\quad A_1 A^{-1} A_2 = R_1^{-1} R_1 (R_1+R_2)^{-1} R_2 R_2^{-1} = (R_1+R_2)^{-1} \qquad (7)$

Using (5),(6), and (7), from (4) the lemma follows.

APPENDIX C

Flow chart of the iteration

Start

IT = 1

Read R_1, R_2, ϕ and initial $\theta = \theta 1$, ε

$\theta 1 = \theta 2$

Calculate $\underset{\sim}{\alpha}$ using (3.3.6)

Calculate $\theta 2$ using (3.3.7)

IS IT $\geq$ ITMAX?

IT=IT+1

IS ABS($\theta 2 - \theta 1$) $\leq$?

NO

NO

Yes

Yes

Write $\underset{\sim}{\alpha}$, $\theta 2$

Stop
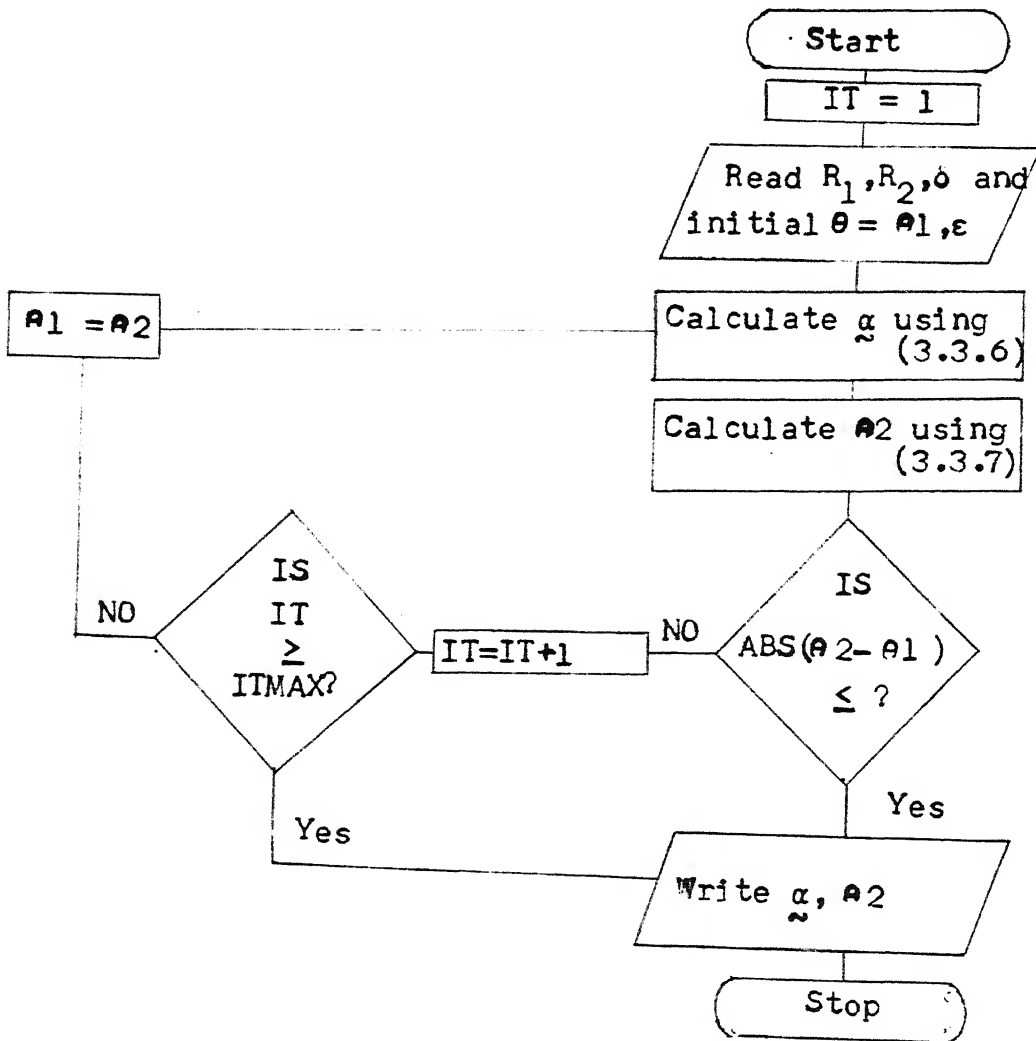
IT   =   number of iteration
ITMAX = maximum number of iteration allowed
ABS   = absolute value
ε = pre-assigned small quantity

APPENDIX D

Solutions of equations by Graeffe's method

Suppose the equation to be solved is

$$f(x) = 0 \qquad\qquad (1)$$

Let $f(x)$ be expressed in the form :

$$f(x) = (x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_n)$$

where

$$f(x) = \sum_{i=1}^{n} a_i x^{n-i} \ , \ a_o = 1$$

Here n is the degrees of the equation, $a_o, a_1, \ldots, a_n$ are the coefficients and $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the roots. It is understood that $a_n \neq 0$. We assume α_i's are real and distinct.

Consider the function φ defined by

$$\varphi(x) = (-1)^n f(x)f(-x)$$
$$= (x^2-\alpha_1^2)(x^2-\alpha_2^2)\ldots(x^2-\alpha_n^2) \qquad (2)$$

Since $\varphi(x)$ is a polynomial containing only even powers, we may define the polynomial

$$f_2(x) = \varphi(\sqrt{x}) = (x-\alpha_1^2)(x-\alpha_2^2)\ldots(x-\alpha_n^2),$$

which has the property that the roots of $f_2(x) = 0$ are the squares of the roots (1). Repeating this operation, we obtain a sequence of polynomials $f_2, f_4, f_8, f_{16}, \ldots$, such that the equation

$$f_m(x) \equiv (x-\alpha_1^m)(x-\alpha_2^m)\ldots(x-\alpha_n^m) = o, \tag{3}$$

where m is a positive integral power of 2, has the roots $\alpha_1^m, \ldots, \alpha_n^m$. If the roots of (1) real and $|\alpha_1| > |\alpha_2| > \ldots > |\alpha_n|$, then the ratios

$$|\alpha_2^m/\alpha_1^m|, \ |\alpha_3^m/\alpha_2^m| \ , \ldots, \ |\alpha_n^m/\alpha_{n-1}^m|$$

can be made as small as desired by making m large enough. Expanding (3) leads to

$$f_m(x) = x^n - (\alpha_1^m + \ldots)x^{n-1} + (\ \alpha_1^m\alpha_2^m + \ldots)x^{n-2}$$
$$-(\alpha_1^m\ \alpha_2^m\ \alpha_3^m + \ldots)x^{n-3} + \ldots + (-1)^n\ \alpha_1^m\alpha_2^m\ldots\alpha_n^m \tag{4}$$

Writing the right-hand side of (4) in the form

$$x^n - A_1 x^{n-1} + A_2 x^{n-2} + \ldots + (-1)^n\ A_n \ ,$$

we derive the approximation

$$\alpha_1^m \doteq A \ , \ \alpha_2^m \doteq \frac{A_2}{A_1} \ ,\ldots, \ \alpha_n^m \doteq \frac{A_n}{A_{n-1}} \tag{5}$$

From these , by taking mth roots, we may approximate the values of the roots $\alpha_1, \ldots, \alpha_n$ of (1).

Since the signs of the roots are not determined, by they must be checked by substitution.

The coefficient may be found iteratively by the relations

$$_{j+1}A_i = (-1)^i \left[_jA_i^2 + 2 \sum_{\ell=1}^{i} (-1)^\ell \ _jA_{i+\ell}\ _jA_{i-\ell} \right], \ 0 \leq i \leq n \tag{6}$$

where $_oA_i = a_i$ and $f_{2j}(x) = \sum_{i=o}^{n}\ _jA_i\ x^{n-i} \ , \ 0 \leq i \leq n$

In the above formula, the presubscripts on A refer to the iteration counter, that is, $_jA_i$ is the value of A_i found

on the jth pass of the iterative scheme ; $_oA_i$ is the initial value of A_i. Indices greater than n mean that the number to be used is zero.

Method of Solution :

Let coefficient $_jA_i$ and $_{j+1}A_i$ described in (6) be rewritten as

$$c_i = {}_jA_i$$
$$B_i = {}_{j+1}A_i.$$

Then, given the initial values for the c_i , where

$$c_i = {}_oA_i = a_i,$$

the B_i for one iteration may be evaluated from

$$B_o = 1$$
$$B_i = (-1)^i\left[\ c_i^2 + 2 \sum_{\ell=1}^{i} (-1)^\ell\ c_{i+\ell}\, c_{i-\ell}\ \right],$$

$$i = 1,2,\dots,n$$

APPENDIX E

Two Lemmas

Lemma : (a) $\lim\limits_{n \to \infty} \dfrac{1}{n} \underset{\sim}{\delta}' \Sigma \underset{\sim}{\delta} = \dfrac{1}{2\pi} \int_{-\pi}^{\pi} h(\lambda) dM(\lambda).$

Proof: $\underset{\sim}{x}' = (x(o), \ldots, x(n-1)$ is covariance stationary

with mean $\quad \underset{\sim}{\mu}_j = (\underset{\sim}{\mu}_j(o), \ldots, \underset{\sim}{\mu}_j(n-1))$, $j = 1, 2,$

and covariance $\quad \Sigma_j = ((\sigma_j(s-t)))$, $s, t = o, \ldots, n-1$

$$\sigma_j(s-t) = Cov(X(s), X(t)), \text{ under } H_j.$$

Define the Discrete Fourier transform of $\{\sigma(n)\}$ sequence :

$$h(\lambda) \equiv \sum_{n=-\infty}^{\infty} \sigma(n) e^{in\lambda}$$

If a spectral density $h(\lambda)$ is continuous, then for an

arbitrary $\varepsilon > o$ there are two trigonometric polynomials

$$h_L(\lambda) = \sum_{k=-K}^{K} \sigma_L(k) e^{i\lambda k} \text{ and } h_U(\lambda) = \sum_{k=-K}^{K} \sigma_i(k) e^{i\lambda k}$$

with $\sigma_L(k) = \sigma_L(-k) \quad$ and $\sigma_U(k) = \sigma_U(-k)$

such that

$$h_L(\lambda) \leq h(\lambda) \leq h_U(\lambda), \quad -\pi \leq \lambda \leq \pi \tag{1}$$

$$h_U(\lambda) - h_L(\lambda) \leq \varepsilon .$$

Define

$$\Sigma_L = ((\sigma_L(s-t))), \; \Sigma_U = ((\sigma_U(s-t))).$$

Now, for any arbitrary vector x,

$$\underset{\sim}{x}' \Sigma \underset{\sim}{x} = \sum_{s,t=o}^{n-1} \sigma(s-t) x_t x_s$$

$$= \int_{-\pi}^{\pi} \sum_{s,t} e^{i\lambda(s-t)} h(\lambda) x_s x_t \frac{d\lambda}{2\pi}$$

$$= \int_{-\pi}^{\pi} \left| \sum_{s=0}^{n-1} e^{i\lambda s} x_s \right|^2 h(\lambda) \frac{d\lambda}{2\pi}$$

Similarly,

$$\underset{\sim}{x}' \not\!\!\Gamma_L \underset{\sim}{x} = \int_{-\pi}^{\pi} \left| \sum_{0}^{n-1} x_s e^{i\lambda s} \right|^2 h_L(\lambda) \frac{d\lambda}{2\pi}$$

$$\underset{\sim}{x}' \not\!\!\Gamma_U \underset{\sim}{x} = \int_{-\pi}^{\pi} \left| \sum_{0}^{n-1} x_s e^{i s} \right|^2 h_U(\lambda) \frac{d\lambda}{2\pi}$$

Then from (1),

$$\int_{-\pi}^{\pi} \left| \sum_{0}^{n-1} x_s e^{i\lambda s} \right|^2 h_L(\lambda) \frac{d\lambda}{2\pi}$$

$$\leq \int_{-\pi}^{\pi} \left| \sum_{0}^{n-1} x_s e^{i\lambda s} \right|^2 h(\lambda) \frac{d\lambda}{2\pi}$$

$$\leq \int_{-\pi}^{\pi} \left| \sum_{0}^{n-1} x_s e^{i\lambda s} \right|^2 h_U(\lambda) \frac{d\lambda}{2\pi}$$

$$\Longrightarrow \quad \underset{\sim}{x}' \not\!\!\Gamma_L \underset{\sim}{x} \leq \underset{\sim}{x}' \not\!\!\Gamma \underset{\sim}{x} \leq \underset{\sim}{x}' \not\!\!\Gamma_U \underset{\sim}{x} ,$$

$$\text{for any vector } \underset{\sim}{x}.$$

Now,

$$\frac{1}{n}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \not\!\!\Gamma (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)$$

$$\leq \frac{1}{n}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \not\!\!\Gamma_U (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)$$

$$= \frac{1}{n} \int_{-\pi}^{\pi} \left| \sum_{t=0}^{n-1} \delta(t) e^{i\lambda t} \right|^2 h_U(\lambda) \frac{d\lambda}{2\pi}$$

$$= \frac{1}{n} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\sum_{t,s=0}^{n-1} \delta(t)(s) e^{i\lambda(t-s)} \right) \left(\sum_{k=-K}^{K} \sigma_U(k) e^{i\lambda k} \right) d\lambda$$

$$= \frac{1}{n.2\pi} \int_{-\pi}^{\pi} \sum_{k=-K}^{K} \sum_{t,s=0}^{n-1} \delta(t)\delta(s) \; \sigma_U(k)e^{i\lambda(t-s+k)} \; d\lambda$$

$$= \frac{1}{2\pi n} . \; 2\pi \sum_{k=-K}^{K} \sigma_U(k) \sum_{t \, \varepsilon \, S_k} \delta(t)\delta(t+k) ,$$

$\left(\text{where} \quad S_k = \{(0,\ldots,n-1-k)\} \text{ if } k \geq 0\right.$

$$\{(-k,\ldots,n-1)\} \quad \text{if } k \leq 0 \; \left.\right)$$

$$= \sum_{k=-K}^{K} \sigma_U(k) \{ \frac{1}{n} \sum_{t=0}^{n-1-|k|} \delta(t)\delta(t+|k|)\}$$

$$\overline{\lim_{n}} \; \frac{1}{n} \underset{\sim}{\delta'} \not{\cancel{}} \underset{\sim}{\delta}$$

$$\leq \sum_{k=-K}^{K} \sigma_U(k) . \; \{\overline{\lim_{n}} \frac{1}{n} \sum_{t=0}^{n-1-|k|} \delta(t)\delta(t+|k|)\}$$

$$= \sum_{k=-K}^{K} \sigma_U(k) \{ \int_{-\pi}^{\pi} e^{i\lambda k} \; \frac{dM(\lambda)}{2\pi} \} , \; \text{(by asymption A3 of Chapter II}$$

$$= \int_{-\pi}^{\pi} (\sum_{k=-K}^{K} \sigma_U(k) \; e^{i\lambda k}) \; \frac{dM(\lambda)}{2\pi}$$

$$= \int_{-\pi}^{\pi} h_U(\lambda) \; \frac{dM(\lambda)}{2\pi}$$

Similarly ,

$$\underline{\lim_{n}} \; \frac{1}{n} \underset{\sim}{\delta'} \not{\cancel{}} \underset{\sim}{\delta} \geq \int_{-\pi}^{\pi} h_L(\lambda) \; \frac{dM(\lambda)}{2\pi}$$

Therefore, $\int_{-\pi}^{\pi} h_L(\lambda) \; \frac{dM(\lambda)}{2\pi} \leq \underline{\lim_{n}} \; \frac{1}{n} \underset{\sim}{\delta'} \not{\cancel{}} \underset{\sim}{\delta}$

$$\leq \overline{\lim_{n}} \; \frac{1}{n} \underset{\sim}{\delta'} \not{\cancel{}} \underset{\sim}{\delta}$$

$$\leq \int_{-\pi}^{\pi} h_U(\lambda) \; \frac{dM(\lambda)}{2\pi}$$

This complete the proof.

(b) $\lim\limits_{n \to \infty} \frac{1}{n} \underset{\sim}{\delta}' R_1 (R_1+R_2)^{-1} R_1 \underset{\sim}{\delta} = \int_{-\pi}^{\pi} \frac{f_1^2(\lambda)}{(f_1+f_2)(\lambda)} \frac{dM(\lambda)}{2\pi}$

Proof : Let R_1^0 be a matrix whose (s,t)th element is

$$r_1^0(s-t) = \frac{1}{n} \sum_{m=0}^{n-1} f_1(\lambda_m) e^{i\lambda_m(s-t)}$$

and R_1^{0-1} be a matrix where (s,t)th element is

$$r^{0-1}(s-t) \cong \frac{1}{n} \sum_{n=0}^{n-1} \frac{1}{f_1(\lambda_m)} e^{i\lambda_m(s-t)}$$

Similarly, we define R_2^0 and R_2^{0-1}.

We have, $\underset{\sim}{\delta}' R_1^0 (R_1^0+R_2^0)^{-1} R_2^0 \underset{\wedge}{\delta} = \underset{\wedge}{\delta}' (R_1^{0-1} + R_2^{0-1})^{-1} \underset{\sim}{\delta}$

(s,t)th element of $R_1^{0-1} + R_2^{0-1} = \frac{1}{n} \sum_{m=0}^{n-1} \frac{f_1(\lambda_m)+f_2(\lambda_m)}{f_1(\lambda_m)f_2(\lambda_m)} e^{i\lambda_m(s-t)}$

Thus, (s,t)th element of $(R_1^{0-1}+R_2^{0-1})^{-1}$

$$= \frac{1}{n} \sum_{m=0}^{n-1} \frac{f_1(\lambda_m)f_2(\lambda_m)}{f_1(\lambda_m)+f_2(\lambda_m)} e^{i\lambda_m(s-t)}$$

Now we can write

$$\underset{\sim}{\delta}' R_1^0 (R_1^0+R_2^0)^{-1} R_2^0 \underset{\sim}{\delta} = \underset{\sim}{\delta}' R_1^0 \underset{\sim}{\delta} - \underset{\sim}{\delta}' R_1^0 (R_1^0+R_2^0)^{-1} R_1^0 \underset{\sim}{\delta}$$

so that

$$\underset{\sim}{\delta}' R_1^0 (R_1^0+R_2^0)^{-1} R_1^0 \underset{\sim}{\delta} = \underset{\sim}{\delta}' R_1^0 \underset{\sim}{\delta} - \underset{\sim}{\delta}' R_1^0 (R_1^0+R_2^0)^{-1} R_2^0 \underset{\sim}{\delta}$$

$$= \sum_{m=0}^{n-1} f_1(\lambda_m) |D(\lambda_m)|^2 - \sum_{m=0}^{n-1} \frac{f_1(\lambda_m)f_2(\lambda_m)}{f_1(\lambda_m)+f_2(\lambda_m)} |D(\lambda_m)|^2$$

$$= \sum_{m=0}^{n-1} \frac{f_1^2(\lambda_m)}{f_1(\lambda_m)+f_2(\lambda_m)} |D(\lambda_m)|^2 \, ,$$

where $D(\lambda_m) = \dfrac{1}{\sqrt{n}} \sum\limits_{t=0}^{n-1} \delta(t)e^{-i\lambda_m t}$ is the Finite Fourier transform of the mean difference function (see [5]) and $\lambda_m = \dfrac{2\pi m}{n}$ $(m = 0,1,..,n-1)$.

We want to show that

i) $\quad \lim\limits_{n \to \infty} \dfrac{1}{n} \left| \underset{\sim}{\delta}' R_1^o ((R_1+R_2)^{-1} - (R_1+R_2)^{o-1}) R_1^o \underset{\sim}{\delta} \right| = 0$

ii) $\quad \lim\limits_{n \to \infty} \dfrac{1}{n} \left| \underset{\sim}{\delta}' R_1 (R_1+R_2)^{-1} R_1 \underset{\sim}{\delta} - \underset{\sim}{\delta}' R_1^o (R_1+R_2)^{-1} R_1^o \underset{\sim}{\delta} \right| = 0$

Combining (i) and (ii), we get

$$\lim\limits_{n \to \infty} \dfrac{1}{n} \left| \underset{\sim}{\delta}' R_1 (R_1+R_2)^{-1} R_1 \underset{\sim}{\delta} - \underset{\sim}{\delta}' R_1^o (R_1+R_2)^{o-1} R_1^o \underset{\sim}{\delta} \right| = 0$$

Hence, $\quad \lim\limits_{n \to \infty} \dfrac{1}{n} \underset{\sim}{\delta}' R_1 (R_1+R_2)^{-1} R_1 \underset{\sim}{\delta}$

$$= \lim\limits_{n \to \infty} \dfrac{1}{n} \underset{\sim}{\delta}' R_1^o (R_1+R_2)^{o-1} R_1^o \underset{\sim}{\delta}$$

$$= \lim\limits_{n \to \infty} \dfrac{1}{n} \sum\limits_{m=0}^{n-1} \dfrac{f_1^2(\lambda_m)}{f_1(\lambda_m)+f_2(\lambda_m)} |D(\lambda_m)|^2$$

$$= \int_{-\pi}^{\pi} \dfrac{f_1^2(\lambda)}{f_1(\lambda)+f_2(\lambda)} \dfrac{dM(\lambda)}{2\pi} ,$$

(see [59]).

Proof of (i) : Step 1 We prove, $\sup\limits_{p} \sum\limits_{t} |r_1^{tp}| < \infty$

where r_1^{tp} is the (t,p)th element of R_1^{-1}.

We have, $||R_1|| = \sup\limits_{||x|| \leq 1} ||R_1 x||$

$$= \sup\limits_{||x|| \leq 1} \max\limits_{i} \left(\left| \sum\limits_{j=1}^{n} r_{1ij} x_j \right| \right)$$

$$\leq \sup_{||x|| \leq 1} \max_i \left(\sum_{j=1}^n |r_{1ij}||x_j| \right)$$

$$\leq \sup_{||x|| \leq 1} \max_i \left(\sum_{j=1}^n |r_{1ij}| \right)$$

$$= \max_i \left(\sum_{j=1}^n |r_{1ij}| \right)$$

Let

$$x_o^{(i)} = \{ \frac{r_{1ij}}{|r_{1ij}|} \}_{j=1}^n$$

$$\Rightarrow ||x_o^{(i)}||_{\ell^\infty} \leq 1 \quad \forall i$$

Thus, $||R_1|| \geq ||R_1 x_o^{(i)}||$

$$= \max_i \left| \sum_j r_{1ij} x_{oj}^{(i)} \right|$$

$$= \max_i \left| \sum_j |r_{1ij}| \right|$$

$$= \max_i \left(\sum_{j=1}^n |r_{1ij}| \right)$$

Hence, $||R_1|| = \max_i \left(\sum_{j=1}^n |r_{1ij}| \right)$

Equivalently, $||R^{-1}|| = \max_p \left(\sum_t |r_1^{tp}| \right)$

$||R_1^{-1}||$ is bounded if $||R_1||$ is bounded away from zero ; this is proved in ([32]).

Step 2 : We shall now show, $\sum_p \sum_q |r_1(p-q) - r_1^0(p-q)| < \infty$

$$r_1^0(p-q) = \frac{1}{n} \sum_{m=0}^{n-1} f(\lambda_m) e^{i\lambda_m(p-q)}$$

$$= \frac{1}{n} \sum_{m=0}^{n-1} f\left(\frac{2\pi m}{n}\right) e^{i\frac{2\pi m}{n}(p-q)}$$

$$= \frac{1}{n} \sum_{m=0}^{n-1} \left\{ \sum_{\tau=-\infty}^{\infty} r_1(\tau) e^{-i\frac{2\pi m}{n}\tau} \right\} e^{i\frac{2\pi m}{n}(p-q)}$$

$$== \frac{1}{n} \sum_{m=0}^{n-1} \sum_{\tau=-\infty}^{\infty} e^{-i\frac{2\pi m}{n}\tau} e^{i\frac{2\pi m}{n}(p-q)} r(\tau)$$

$$= \sum_{\ell=-\infty}^{\infty} r_1(p-q+\ell n),$$

(put $p-q-\tau = -\ell n$, $\ell = 0, \pm 1, \pm 2, \ldots$).

Thus,
$$\sum_{p,q=0}^{n-1} |r_1(p-q) - r_1^0(p-q)|$$

$$= \sum_{p,q=0}^{n-1} \left| r_1(p-q) - \sum_{\ell=-\infty}^{\infty} r_1(p-q+\ell n) \right|$$

$$= \sum_{\tau=-(n-1)}^{n-1} (n-|\tau|) \left| r_1(\tau) - \sum_{\ell=-\infty}^{\infty} r_1(\tau+\ell n) \right|$$

$$= \sum_{\tau=-(n-1)}^{n-1} (n-|\tau|) \left| \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} r_1(\tau+\ell n) \right|$$

$$\leq \sum_{\tau=-(n-1)}^{n-1} (n-|\tau|) \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} |r_1(\tau+\ell n)|$$

$$= \sum_{\tau=-(n-1)}^{n-1} (n-|\tau|) \left\{ \sum_{\substack{\ell=-\infty \\ \ell \neq 0, \pm 1}}^{\infty} |r_1(\tau+\ell n)| + |r_1(\tau+n)| + |r_1(\tau-n)| \right\}$$

$$= \sum_{\tau=-(n+1)}^{n-1} (n-|\tau|) \{ \sum_{\substack{\ell=-\infty \\ \ell \neq 0, \pm 1}}^{\infty} |r_1(\tau+\ell n)| \}$$

$$+ \sum_{\tau=-(n-1)}^{n-1} (n-|\tau|) |r_1(\tau+n)|$$

$$+ \sum_{\tau=-(n-1)}^{n-1} (n-|\tau|) |r_1(\tau-n)|$$

$$\leq n \sum_{\substack{\ell=-\infty \\ \ell \neq 0, \pm 1}}^{\infty} \sum_{\tau=-(n-1)}^{n-1} |r_1(\tau+\ell n)|$$

$$+ n \sum_{0}^{n-1} |r_1(\tau+n)| + n \sum_{-(n-1)}^{0} |r_1(\tau-n)|$$

$$+ \sum_{-(n-1)}^{-1} (n-|\tau|) |r_1(\tau+n)| + \sum_{1}^{n-1} (n-|\tau|) |r_1(\tau-n)|$$

$$\leq 2 \sum_{|v| \geq n} |v| |r_1(v)|$$

Step 3 : We prove,

$$\sum_{t,t'} |r_1^{tt'} - r_1^{ott'}|$$

$$\leq (\sup_{p} \sum_{t} |r_1^{tp}|)(\sup_{q} \sum_{t'} |r_1^{oqt'}|) [\sum_{pq} \sum |r_1(p-q) - r_1^o(p-q)|]$$

Since for any non-singular matrices A and B, $A^{-1}(A-B)B^{-1} = B^{-1} - A^{-1}$,

$$|r_1^{ott'} - r_1^{tt'}| = |(R_1^{o-1} - R_1^{-1})_{t,t'}|$$

$$= |[r_1^{t1}, \ldots, r_1^{tn}](R_1 - R_1^o) \begin{bmatrix} r_1^{olt'} \\ \cdot \\ \cdot \\ r_1^{ont'} \end{bmatrix} |$$

$$= | \sum_{k,j} r_1^{tk} (r_1(k-j)-r_1^o(k-j))r_1^{ojt'} |$$

$$\le (\sup_p \sum_t |r_1^{tp}|)(\sup_q \sum_{t'} |r^{oqt'}|)(\sum_{p,q} |r_1(p-q)-r_1^o(p-q)|)$$

Step 4 : Proof of (i) will be complete if we observe, following the above steps together with ([59]) that

$$| \underset{\sim}{\delta}' R_1^o((R_1+R_2)^{-1}-(R_1+R_2)^{o-1})R_1^o \underset{\sim}{\delta} |$$

$$\le (\sup_t |\delta(t)|)^2(\sup_t \sum_{t'} |r^o(t-t')|)^2(\sup_p \sum_t |r^{tp}|)(\sup_q \sum_{t'} |r^{oqt'}|)$$

$$\times [\sum_p \sum_q |r(p-q)-r^o(p-q)|]$$

where $r(p-q) = (p,q)$th element of (R_1+R_2).

Proof of (ii) : It follows immediately (see [59]) if we write,

$$| \underset{\sim}{\delta}' R_1 (R_1+R_2)^{-1}R_1\underset{\sim}{\delta} - \underset{\sim}{\delta}' R_1^o(R_1+R_2)^{-1}R_1^o \underset{\sim}{\delta}|$$

$$\le [\sup_s | \sum_{u,t} \delta(u)r_1(u-t)r^{-1}(s-t)|$$

$$+ \sup_t | \sum_{u,s} r^{-1}(t-u)r_1^o(u-s)\delta(s)|](\sum_{s,t} |r_1(s-t)-r_1^o(s-t)|).$$

APPENDIX F

Series Representation of a stochastic process

<u>Theorem</u> : Write $x(t) = \sum\limits_{n=o}^{\infty} x_n \varphi_n(t)$.

Then $\qquad E_{H_j}(x(t)-x(t))^2 = o.$

And if $\qquad r_j(t,u) = \sum \beta_{jn}(u) \varphi_n(t)$ (for a given u),

where $\qquad \beta_{jn}(u) = \int_o^A r_j(t,u) \varphi_n(t) dt$

and $\qquad \int_o^A \varphi_n(t) \varphi_m(t) dt = \delta_{nm}$,

then $\qquad E_{H_j} x_n x_m = \int_o^A \beta_{jn}(u) \varphi_m(u) du$ (j = 1,2)

provided $Ex^2(t) < \infty.$

<u>Proof</u> : $E_{H_j}(x(t)-x(t))^2$

$\qquad = E_{H_j}x^2(t) - 2E_{H_j}x(t) \sum\limits_{n=o}^{\infty} x_n \varphi_n(t) + E \sum\limits_{n,m} x_n x_m \varphi_n(t)\varphi_m(t)$

Now, $\qquad E_{H_j}x(t) \sum\limits_{o}^{\infty} x_n \varphi_n(t)$

$\qquad = E \sum\limits_{n} x(t)(\int_o^A x(u)\varphi_n(u)du) \varphi_n(t)$

$\qquad = \sum\limits_{o}^{\infty} (\int_o^A r_j(t,u) \varphi_n(u)du) \varphi_n(t)$

$\qquad = \sum\limits_{n=o}^{\infty} \beta_{jn}(t) \varphi_n(t) = r_j(t,t)$

$$E \sum_{n,m=o}^{\infty} x_n x_m \, \varphi_n(t) \, \varphi_n(t)$$

$$= \Sigma (Ex_n x_m) \, \varphi_n(t) \, \varphi_n(t)$$

$$= \sum_{n,m} (\int_o^A \beta_{jn}(u) \, \varphi_m(u)du) \, \varphi_n(t) \, \varphi_m(t),$$

(s ince

$$E \, x_n x_m = \int_o^A (\int_o^A r_j(t,u) \, \varphi_n(t)dt) \, \varphi_m(u)du = \int_o^A \beta_{jn}(u)\varphi_m(u)du$$

$$= \sum_{m=o}^{\infty} \{ \int_o^A (\sum_{n=o}^{\infty} \beta_{jn}(u)\varphi_n(t))\varphi_m(u)du \} \, \varphi_m(t)$$

$$= \sum_{m=o}^{\infty} \{ \int_o^A r_j(t,u) \, \varphi_m(u)du \} \, \varphi_m(t)$$

$$= \sum_{m=o}^{\infty} \beta_{jm}(t) \, \varphi_m(t)$$

$$= r_j(t,t).$$

This completes the proof.

A complex process $\{Z(t), \, t\epsilon T\}$ has the similar representation

$$Z(t) = \sum_{n=o}^{\infty} z_n \, \varphi_n(t)$$

where $\quad Z_n = {}_o\int^A z(t) \, \varphi_n(t)dt \, , \, {}_o\int^A \varphi_n(t) \, \overline{\varphi_m(t)}dt = \delta_{nm}$

For, proof, follow the above approach and note that in this

case,

$$R(t,u) = \Sigma \, \beta_n(n) \, \varphi_n(t)$$

$$R(u,t) = \overline{R(t,u)} = \Sigma \, \overline{\beta_n(u)} \, \overline{\varphi_n(t)}$$

i.e. $\quad R(t,u) = \sum_n \overline{\beta_n(t)} \, \overline{\varphi_n(u)}$

APPENDIX G

The Dirac Delta function

Suppose that $\varphi(t)$ is any function which is continuous at $t = 0$. Then the Dirac delta function $\delta_D(t)$ is such that

$$\int_{-\infty}^{\infty} \delta_D(t)\, \varphi(t)dt = \varphi(o) \qquad (1)$$

(see [14]). It is important to realize that $\delta_D(t)$ is not a function. Rather it is a generalized function which maps a function into the real line.

Even though $\delta_D(t)$ is a generalized function, it can often be handled as if it were an ordinary function except that we will be interested in the value of the integral involving $\delta_D(t)$ and never in the value of $\delta_D(t)$ by itself.

The derivative $\delta_D'(t)$ of $\delta_D(t)$ can also be defined by

$$\int_{-\infty}^{\infty} \delta_D'(t)\, \varphi(t)dt = -\varphi'(o)$$

where $\varphi'(o)$ is the derivative of $\varphi(t)$ evaluated at $t=0$. The justification for the above depends on integrating by parts as if $\delta_D'(t)$ and $\delta_D(t)$ were ordinary functions and using the following definition of delta function (which is heuristically useful),

$$\delta_D(t) = \begin{cases} o & t \neq o \\ \infty & t = o \end{cases}$$

such that $\int_{-\infty}^{\infty} \delta(t)dt = 1.$

We have

$$\int_{-\infty}^{\infty} \delta_D'(t) \; \varphi(t)dt = - \int_{-\infty}^{\infty} \delta_D(t) \; \varphi'(t)dt = - \varphi'(o)$$

APPENDIX H

Differentiation with respect to a matrix

Let $Z = ((z_{jk}))$ be an nxm complex matrix. Let g be a scalar valued function of Z. If we write $z_{jk} = x_{jk} + iy_{jk}$ where $x_{jk} = \text{Re } z_{jk}$, and $y_{jk} = \text{Im } z_{jk}$, then g may be considered to be a function of 2mn variables, x_{jk}, y_{jk}, $1 \leq j \leq n$, $1 \leq k \leq m$. Now suppose g is a differentiable function of these 2mn variable in some region of 2mn dimensional real Euclidean space. Then we define (see [40]) the derivative of g with respect to Z as the nxm matrix

$$\frac{dg}{dZ} = \frac{1}{2}((\frac{\partial g}{\partial x_{jk}} - i\frac{\partial g}{\partial y_{jk}})).$$

For example, if ξ and c are n-dimensional complex vectors, then a direct application of this definition yields,

$$\frac{d}{d\xi}(c'\xi) = c \tag{1}$$

and

$$\frac{d}{d\xi}(c'\bar{\xi}) = o \tag{2}$$

If Q is an arbitrary nxn complex matrix, then

$$\frac{d}{d\xi}(\xi'Q\xi) = (Q+Q')\xi \tag{3}$$

$$\frac{d}{d\xi}(\bar{\xi}'Q\xi) = Q'\bar{\xi} \tag{4}$$

$$\frac{d}{d\xi}(\xi'Q\bar{\xi}) = Q\bar{\xi} \tag{5}$$

APPENDIX I

Integration of Complex Stochastic Processes

Let $\{Z(t), t \in T\}$ be a complex stochastic process
defined on a probability space $(\Omega, \mathscr{F}, P)$.

Let $T = [a,b]$ be a closed finite interval.

Let n be an integer and

$$a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b$$

be a partition, π, of T. In each sub-interval $[t_{k-1}, t_k]$
choose an arbitrary point and call it t_k'. The partition π
now becomes a marked partition, π'.

Define

$$S_{\pi'} = \sum_{k=1}^{n} Z(t_k')(t_{k-1} - t_k)$$

which, called the approximating sum, is a complex random
variable.

If $S_{\pi'}$ converges in mean square as the norm
$\nu (\overset{\Delta}{=} \max_{1 \leq k \leq n} (t_k - t_{k-1}))$ of the partition π approaches zero,
the limit will be called the _mean square integral_ of
$\{Z(t), t \in T\}$ and we shall write

$$\lim_{\nu \to 0} S_{\pi'} = \int_T Z(t)\, dt$$

We shall give some sufficient conditions under which the
integral of a complex stochastic process exists.

Theorem (Miller [40], p.105) : Let $\{Z(t), t \in T\}$ be a
complex stochastic process with mean zero and covariance

function $R(t,s) = EZ(t)\overline{Z(s)}$. Let R be continuous on TxT.
Then Z(t) is mean square integrable on T (which is a finite
interval [a,b]) i.e.

$$\int_T Z(t)\ dt \text{ exists in m.s.}$$

Theorem (Miller [40], p. 109) : Let T [a,b] be a
closed interval and let $\{Z(t), t \varepsilon T\}$ be a complex stochastic
process with mean zero and covariance function $R(t,s)$.
Let R be continuous on TxT and let g be a complex valued
function of the real variable t which is continuous on T.
Then

$$\int_a^b g(t)Z(t)dt$$

exists as a mean square integral.

Now, we shall see how we define integral of a
complex process when $T = [0,\infty)$

Theorem (Miller [40], p.14) : Let $\{Z(t), 0 < t < \infty\}$
be a complex process with mean zero and covariance function
$R(t,s) = E\ Z(t)\ \overline{Z(s)}$. Let R be continuous on $[0,\infty)\times[0,\infty)$
and let g be a complex-valued function of the real variable
t which is continuous on $[0,\infty)$. Let

$$\lim_{b,b' \to \infty} \int_0^b \int_0^{b'} g(t)\overline{g(s)}\ R(t,s)dtds$$

exists. Then

$$\xi \overset{\Delta}{=} \int_0^\infty g(t)Z(t)dt$$

exists as a mean square integral and

$$E\xi = o$$

$$E|\xi|^2 = \int_o^\infty \int_o^\infty g(t)\,\overline{g(s)}R(t,s)\,dt\,ds$$

REFERENCES

[1]   Adhikari, B.P. and Joshi, D.D., ''Distance, discrimination et re'sume' exhaustif'', Publ. Inst. Statist. Univ., Paris, 5, 1956, pp. 57-74.

[2]   Ali, S.M. and Silvey, S.D., ''A General class of Coefficients of Divergence of One Distribution from Another'', Journal of the Royal Statistical Society, Ser B, 28, 1966, pp. 131-142.

[3]   Anderson, T.W., ''An Introduction to Multivariate Statistical Analysis'', John Wiley and Sons, Inc., 1958.

[4]   Anderson, T.W. and Bahadur, R.R., ''Classification into two Multivariate Normal distributions with Different Covariance matrices'', Annals of Mathematical Statistics, 33, (June, 1962), pp. 420-431.

[5]   Anderson, T.W., ''The Statistical Analysis of Time Series'', Wiley, New York, 1971.

[6]   Apostol, T.M., ''Mathematical Analysis'', 2nd Edition, Addison-Wesley Publishing Company, Massachusetts, 1981.

[7]   Ash, R.B. and Gardner, M.F., ''Topics in Stochastic Processes'', Academic Press, New York, 1975.

[8]   Barnard, M.M., ''The secular variation of Skull characters in four series of Egyptian Skulls'', Ann. of Eng., 6, 1935, pp. 352-371.

[9]   Bartlett, M.S. and Please, N.W., ''Discrimination in the case of zero mean differences'', Biometrika, 50, 1963, pp. 17-21.

[10]   Bhat, N., ''Elements of Applied Stochastic Processes'',Ist/ 2nd Edition, John Wiley and Sons, 1984.

[11]   Bhattacharyya, A., ''On a measure of divergence between two statistical populations defined by probability distributions'', Bull. Calcutta Math. Soc., 35, 1943, pp. 99-109.

[12]   Box, G.E.P. and Jenkins, G.M., ''Time Series analysis: Forecasting and Centrol'', Holden-Day, San Francisco, California, 1970.

[13] Cavalli, L.L., ''Alumni prolemi della analisi biometrica di popolazioni naturali '', Mem. Ist. Idrobiol.,2,1945, pp. 301-323.

[14] Chatfield, C., ''The analysis of Time Series'', 2nd Edition, Chapman and Hall, London, 1980.

[15] Chernoff, H., ''A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations'', Ann. Math. Statist., 23, 1952, pp. 493-507.

[16] Clunies-Ross, C.W. and Riffenburgh, R.H., ''Linear discriminant analysis'', Pacif. Sci., 14, 1960, pp. 251-256.

[17] Fisher, R.A., ''The use of Multiple Measurements in taxonomic problems'', Ann. Eng., 7, 1936, pp. 179-188.

[18] Fuller, W.A., ''Introduction to Statistical Time Series'', Wiley, New York, 1976.

[19] Gilbert, E.S., ''The effect of unequal variance-covariance matrices on Fisher's linear discriminant function'', Biometrics. 25, 1969, pp. 505-516.

[20] Gill, P.E. and Murray, W., (Editors), ''Numerical Methods for Constrained optimization'', Academic Press, London, 1974.

[21] Giri, N.C., ''Multivariate Statistical Inference'', Academic Press, New York, 1977.

[22] Grettenburg, T.L., ''Signal Selection in Communication and Radar Systems'', IEEE Trans. IT-9, Oct. 1963, pp. 265-275.

[23] Han, C.P., ''Distribution of discriminant function when covariance matrices are proportional'', Ann. Math. Statist., 40, 1969, pp. 979-985.

[24] Han, C.P., ''Distribution of discriminant function in circular models'', Ann. Inst. Statist. Math., 22, 1970, pp. 117-125.

[25] Hellinger, E., ''Neue begrundung der thorie quadratischer formen von unendlichvielen veränderlichen'', J. für die Reine und angew Math., 36, 1909, pp. 210-271.

[26] Kailath, T., ''The Divergence and Bhattacharyya distance Measures in Signal Selection'', IEEE Trans. on Communication Technology, COM-15 (No.1), 1967, pp. 52-60.

[27] Kakutani, S., ''On equivalance of infinite product
      measures'', Ann. Math. Statist., 49, 1948, pp.214-224.

[28] Kolmogorov, A.N., ''On the approximation of distributions
      of sums of independent summands by infinitely divisible
      distributions'', Sankhya, 25, 1963, pp. 159-174.

[29] Koopmans, L.H., ''The Spectral Analysis of Time Series'',
      Academic Press, New York, 1974.

[30] Kullback, S., ''An application of information theory
      to multivariate analysis'', Ann. Math. Statist., 23,
      1952, pp. 88-102.

[31] Kullback, S., ''Information Theory and Statistics'',
      Wiley, New York, 1959.

[32] Liggett, W.S., ''On the asymptotic optimality of
      spectral analysis for testing hypotheses about time
      series'', Annals of Mathematical Statistics, 42, 1971,
      pp. 1348-1358.

[33] Macon, N., ''Numerical Analysis'' Wiley, New York,
      1963.

[34] Mahalanobis, P.C., ''On the Generalized Distance in
      Statistics'', Proc. Natl. Inst. Science, India ;2,
      1936, pp. 49-55.

[35] Martin, E.S., ''A study of the Egyptian series of
      mandibles with special reference to mathematical
      method of sexing'', Biometrika, 28, 1936, pp. 149-178.

[36] Matusita, K., ''A distance and related statistics
      in Multivariate Analysis'', Proceedings of the
      International Symposium on Multivariate Analysis,
      Ed. P.R. Krishnaiah, New York, Academic Press, 1966,
      pp. 187-200.

[37] Matusita, K., ''On the notion of affinity of several
      distributions and some of its applications'', Ann.
      Inst. Statist. Math., 19, 1967, pp. 181-192.

[38] Matusita, K., ''Classification based on distance in
      multivariate Gaussian case'', Proc. 5th Burkeley Symp.
      Math. Stat. Prob., 1, University of California Press,
      Burkeley, 1967, pp. 299-304.

[39] Matusita, K., ''Some properties of Affinity and
      Applications'', Annals of the Instituto of Statistical
      Mathematics, 23, 1971, pp. 137-155.

[40] Miller, K.S., ''Complex Stochastic Processes'',
      Addison-Wesley Publishing Company, Inc., London, 1974.

[41] Neyman, J. and Pearson, E.S., ''Contribution to the
      theory of Statistical Hypotheses'', Statist. Res. Memo.,
      1, 1936, pp. 1-37.

[42] Okamoto, M., ''An asymptotic expansion for the distribution
      of linear discriminant function'', Ann. Math. Statist.,
      34, 1963, pp. 1286-1301, (Correction: Ann. Math. Statist.,39,
      1968, pp. 1358-1359).

[43] Papoulis, A., ''Probability, Random Variables and
      Stochastic Processes'', 2nd edition, McGraw-Hill Book
      Company, New Delhi, 1984.

[44] Parzen, E., ''Stochastic Processes'', Holden-Day, San
      Francisco, 1962.

[45] Patnaik, P.B., ''The non-central χ^2 and F-distributions
      and their approximation'', Biometrika, 36, 1949,
      pp. 202-232.

[46] Pearson, K., ''On the coefficients of racial likeness',
      Biometrika, 18, 1926, pp. 105-117.

[47] Penrose, L.S. ''Some notes on discrimination'', Ann. Eug.,
      13, 1947, pp. 228-237.

[48] Prasad, S., ''Design of Signals for Communication Systems'',
      M. Tech. Thesis, Indian Institute of Technology, Delhi,
      1971.

[49] Rao, C.R., ''Tests with discriminant functions in
      multivariate analysis'', Sankhya, 7, 1946, pp. 407-413.

[50] Rao, C.R., ''The problem of classification and distance
      between two populations'', Nature (London), 159, 1947a,
      pp. 30-31.

[51] Rao, C.R., ''Statistical criterion to determine the
      group to which an individual belongs'', Nature, 160, 1947b,
      pp. 835-836.

[52] Rao, C.R., ''The utilization of multiple measurements
      in problems of biological classification'', Jour. Roy.
      Statist. Soc., B, 10, 1948, pp. 159-203.

[53] Rao, C.R., ''On the distance between two populations'',
      Sankhya, 9, 1949a, pp. 246-248.

[54] Rao, C.R., ''On some problems arising out of discrimination
      with multiple characters'', Sankhya, 9, 1949b, pp. 343-366.

[55] Rao, C.R., ''Statistical Inference applied to classification
      problems'', Sankhya, 10, 1950, pp. 229-256.

[56] Rao, C.R., ''Advanced Statistical methods in Biometric
      Research'', New York, Wiley, 1952.

[57] Rao, C.R., ''Linear Statistical Inference and its
      applications'', 2nd Edition, Wiley Eastern Private
      Limited, New Delhi, 1973.

[58] Seber, G.A.F., ''Multivariate Observations'', Wiley,
      New York, 1984.

[59] Shumway, R.H. and Unger, A.N., ''Linear Discriminant
      Functions for Stationary time series'', Journal of the
      American Statistical Association, 69, (December, 1974),
      pp. 948-956.

[60] Singh, S., ''Design and Analysis of Experiments for
      Model Discrimination in Uniresponse and Multiresponse
      Systems'', Ph.D. thesis, I.I.T. Kanpur, India, 1986.

[61] Smith, C.A.B., ''Some examples of discrimination'',
      Ann. Eug., 13, 1947, pp. 272-282.

[62] Tricomi, F.G., ''Integral Equations'', Interscience
      Publishers, Inc., New York, 1970.

[63] Van Trees, H.L., ''Detection, Estimation and Modulation
      Theory'', Part I, Wiley, New York, 1968.

[64] Van Trees, H.L., ''Detection, Estimation and Modulation
      Theory'', Part III, Wiley, New York, 1968.

[65] Vilenkin, N. Ya., ''Method of Successive Approximation'',
      Mir Publishers, Moscow, 1979.

[66] Wald, A., ''On a statistical problem arising in the
      classification of an individual into one of two groups'',
      Ann. Math. Statist., 15, 1944, pp. 145-162.

[67]   Wald, A., ''Statistical Decision Function'', Wiley,
       New York, 1950.

[68]   Wald, A. and Wolfowitz, J., ''Characterization
       of minimum complete class of decision function when
       the number of decision is finite, Proc'.' Berkeley
       Symp. Prob. Statist., 2nd, California, 1950.

[69]   Welch, B.L., ''Note on discriminant functions'',
       Biometrika, 31, 1939, pp. 218-220.